(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2003－150419
（P2003－150419A）
(43)公開日　平成15年５月23日(2003.5.23)

| (51)Int.Cl.[7] | | 識別記号 | FI | | テーマコード(参考) |
|---|---|---|---|---|---|
| Ｇ０６Ｆ | 12/00 | ５１４ | Ｇ０６Ｆ 12/00 | ５１４Ｍ | ５Ｂ００５ |
| | | ５１３ | | ５１３Ｄ | ５Ｂ０７５ |
| | 12/08 | ５０５ | 12/08 | ５０５Ｚ | ５Ｂ０８２ |
| | 17/30 | １３０ | 17/30 | １３０Ｚ | |
| | | １５０ | | １５０Ａ | |

審査請求　未請求　請求項の数18　ＯＬ　（全 46 頁）

(21)出願番号　特願2001－348168(P2001－348168)

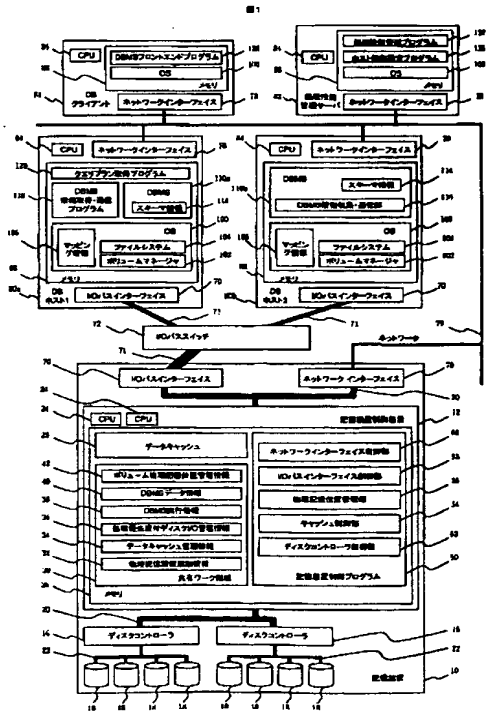
(22)出願日　平成13年11月14日(2001.11.14)

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　茂木　和彦
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(72)発明者　大枝　高
神奈川県川崎市麻生区王禅寺1099番地　株式会社日立製作所システム開発研究所内
(74)代理人　100075096
弁理士　作田　康夫

最終頁に続く

(54)【発明の名称】　データベース管理システムの実行情報を取得する手段を有する記憶装置

(57)【要約】
【課題】記憶装置においてデータベース管理システム（ＤＢＭＳ）の実行情報やデータベースの処理優先度を加味した制御を行うことにより、好適なデータアクセス性能を実現する。
【解決手段】記憶装置は、ＤＢＭＳの静的な構成情報をＤＢＭＳ情報取得・通信プログラム、ＤＢＭＳ情報通信部、ホスト情報設定プログラムを通して取得し、ＤＢＭＳの実行情報をクエリプラン取得プログラム、ＤＢＭＳ情報通信部、処理性能管理プログラムを通して取得し、処理性能管理プログラムから与えられるデータベースの処理の優先度に関する情報を取得し、これらを処理優先度付ディスクＩ／Ｏ管理情報、ＤＢＭＳ実行情報、ＤＢＭＳデータ情報中に記憶する。記憶装置制御プログラム内のキャッシュ制御部はこれらの情報を参照してデータキャッシュの制御を行う。



【特許請求の範囲】

【請求項１】データベース管理システムが稼動している計算機との接続手段を有し、
前記データベース管理システムにおけるスキーマにより定義される表・索引・ログを含むデータ構造に関する情報と、前記データベース管理システムが管理するデータベースデータを前記スキーマにより定義されるデータ構造毎に分類した前記記憶装置における記録位置に関する情報と、前記データベース管理システムが実行する問い合わせ処理の処理実行計画を含む前記データベース管理システムにより管理されるデータベースに関する情報を取得する情報取得手段を有することを特徴とする記憶装置。

【請求項２】前記接続手段を用いて複数の前記データベース管理システムが稼動している計算機と接続することを特徴とする請求項１に記載の記憶装置。

【請求項３】前記情報取得手段が複数の前記データベース管理システムが管理するデータベースに関する情報を取得することを特徴とする請求項１に記載の記憶装置。

【請求項４】前記情報取得手段は前記接続手段を用いて情報を取得することを特徴とする請求項１に記載の記憶装置。

【請求項５】前記情報取得手段が、前記データベース管理システムが管理するデータベースに関する情報を前記データベース管理システムから取得することを特徴とする請求項１に記載の記憶装置。

【請求項６】前記情報取得手段が、前記データベース管理システムが管理するデータベースに関する情報を前記データベース管理システムとは異なる少なくても１つのプログラムを通して取得することを特徴とする請求項１に記載の記憶装置。

【請求項７】前記記憶装置は少なくとも１つ以上のデータを記憶する物理記憶手段を有し、
前記記憶装置内の前記物理記憶手段に対するアクセス順を決定する物理記憶制御手段を有し、
前記物理記憶制御手段が前記物理記憶手段に対するアクセス順を決定する際に前記情報取得手段により取得した情報を利用することを特徴とする請求項１に記載の記憶装置。

【請求項８】キャッシュメモリを有し、
前記キャッシュメモリの管理を行うキャッシュメモリ制御手段を有し、
前記情報取得手段により取得した情報から前記物理記憶手段に対してこれから行われるアクセス先とアクセス方法を予測するアクセス予測手段を有することを特徴とする請求項７に記載の記憶装置。

【請求項９】前記キャッシュメモリ制御手段において、前記アクセス予測手段を利用したキャッシュの破棄データ選択制御を実施することを特徴とする請求項８に記載の記憶装置。

【請求項１０】前記キャッシュメモリ制御手段において、前記アクセス予測手段を利用したデータプリフェッチ実行制御を実施することを特徴とする請求項８に記載の記憶装置。

【請求項１１】前記情報取得手段が取得する情報として、前記データベース管理システムが前記スキーマにより定義される同一のデータ構造に属する前記データベースデータをアクセスする際の並列度に関する情報を含むことを特徴とする請求項１０に記載の記憶装置。

【請求項１２】前記データベースデータの中身を理解するデータベースデータ解釈手段を有し、
前記アクセス予測手段において前記データベースデータ解釈手段を利用することを特徴とする請求項１０に記載の記憶装置。

【請求項１３】前記情報取得手段が取得する情報に、前記データベース管理システムにより管理されるデータベースに与えられた、もしくは実行される処理毎に与えられた処理の優先度に関する処理優先度情報を含むことを特徴とする請求項７に記載の記憶装置。

【請求項１４】前記物理記憶制御手段が前記物理記憶手段に対するアクセス順を決定する際に、前記処理優先度情報を参照して優先度が高い前記データベースのデータを保持する、もしくは高優先度が与えられた処理が利用する前記データベースデータへのアクセスを優先的に実施する制御を行うことを特徴とする請求項１３に記載の記憶装置。

【請求項１５】キャッシュメモリを有し、
前記情報取得手段により取得した情報から前記物理記憶手段に対してこれから行われるアクセス先とアクセス方法を予測するアクセス予測手段を有し、
前記アクセス予測手段を利用したデータプリフェッチを実行し、前記処理優先度情報を加味したキャッシュメモリ制御を実施するキャッシュメモリ制御手段を有することを特徴とする請求項１３に記載の記憶装置。

【請求項１６】前記キャッシュメモリ制御手段において、前記処理優先度情報を参照して優先度が高い前記データベースのデータを保持する、もしくは高優先度が与えられた処理が利用する前記データベースデータに対して優先的にプリフェッチを実施することを特徴とする請求項１５に記載の記憶装置。

【請求項１７】前記キャッシュメモリ制御手段において、前記処理優先度情報を参照して優先度が高い前記データベースのデータを保持する、もしくは高優先度が与えられた処理が利用する前記データベースデータに対してキャッシュメモリ利用量を多く割り当てる制御を実施することを特徴とする請求項１５に記載の記憶装置。

【請求項１８】前記記憶装置内の物理記憶手段の稼動情報を取得する物理記憶稼動情報取得手段を有し、
前記キャッシュ制御手段において、データプリフェッチを実行する際に前記物理記憶稼動情報取得手段により取

得した情報を利用した制御を実施することを特徴とする請求項１５に記載の記憶装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は，データベース管理システムに関する，

【０００２】

【従来の技術】現在、データベース（ＤＢ）を基盤とする多くのアプリケーションが存在し、ＤＢに関する一連の処理・管理を行うソフトウェアであるデータベース管理システム（ＤＢＭＳ）は極めて重要なものとなっている。特に、ＤＢＭＳの処理性能はＤＢを利用するアプリケーションの性能も決定するため、ＤＢＭＳの処理性能の向上は極めて重要である。

【０００３】ＤＢの特徴の１つは、多大な量のデータを扱うことである。そのため、ＤＢＭＳの実行環境の多くにおいては、ＤＢＭＳが実行される計算機に対して大容量の記憶装置を接続し、記憶装置上にＤＢのデータを記憶する。そのため、ＤＢに関する処理を行う場合に、記憶装置に対してアクセスが発生し、記憶装置におけるデータアクセス性能がＤＢＭＳの性能を大きく左右する。そのため、ＤＢＭＳが稼動するシステムにおいて、記憶装置におけるアクセスの最適化が極めて重要である。

【０００４】米国特許５３１７７２７（文献１）においては、無駄なアクセスの削減や必要なデータの先読みによりＤＢＭＳの性能を向上させる技術について開示している。ユーザからの問い合わせ処理（クエリ）を実行する部分において、問い合わせの実行プランやデータアクセス特性、キャッシュメモリ量、Ｉ／Ｏ負荷等を考慮して、プリフェッチ実行やその量の決定、キャッシュ（バッファ）管理等を行うことによりＩ／Ｏアクセス性能を向上させてＤＢＭＳの性能を向上させる。

【０００５】特開平９－２７４５４４号公報（文献２）においては、計算機がアクセスするために利用する論理的記憶装置を実際にデータを記憶する物理的記憶装置上に配置する記憶装置において、前記論理的記憶装置の物理的記憶装置への配置を動的に変更することにより記憶装置のアクセス性能を向上する技術について開示している。アクセス頻度が高い物理記憶装置に記憶されているデータの一部を前記の配置動的変更機能を用いて他の物理記憶装置に移動することにより、特定の物理記憶装置のアクセス頻度が高くならないようにし、これにより記憶装置を全体としてみたときの性能を向上させる。また、配置動的変更機能による高性能化処理の自動実行方法についても開示している。

【０００６】論文“高機能ディスクにおけるアクセスプランを用いたプリフェッチ機構に関する評価”（向井他，第１１回データ工学ワークショップ（ＤＥＷＳ２０００）論文集講演番号３Ｂ－３，２０００年７月発行ＣＤ－ＲＯＭ，主催：電子情報通信学会データ工学研究専門委員会）（文献３）では、記憶装置の高機能化によるＤＢＭＳの性能向上について、リレーショナルデータベース管理システム（ＲＤＢＭＳ）を用いたＤＢを例にして論じている。

【０００７】記憶装置に対してアプリケーションレベルの知識としてＲＤＢＭＳにおける問い合わせ処理の実行時の処理の実行プランを与えた場合、記憶装置は、ＲＤＢＭＳのある表に対する索引を読んだ後、その表のデータを記憶するどのブロックにアクセスすべきなのかが判断できるようになる。そこで、索引を連続的にアクセスし、その索引によりアクセスすべき表のデータを保持するブロック群を把握し、それらに対するアクセスを効果的にスケジューリングすることによりデータの総アクセス時間を短縮することができる。

【０００８】また、この処理はＤＢＭＳが実行されている計算機から独立に実行可能であり、計算機からの命令を待つ必要がない。また、データが複数台の物理記憶装置に分散配置されている場合にはそれぞれの物理記憶装置を並列にアクセス可能であり、よりＤＢＭＳの処理実行時間の短縮が期待できる。

【０００９】文献３においては、前述の効果を擬似実験により確認している。擬似実験においては、実際に記憶装置上に前述の機能を実装するのではなく、ホスト側から先読みの指示を出す方法を採用している。実験に用いた記憶装置は２つのＳＣＳＩポートを保持し、記憶装置内のデータキャッシュは双ポートで共有している。そこで、片方でＤＢＭＳが実際に処理するデータをアクセスし、もう片方のポートで先読みするブロックの読み出しをすることにより、データの先読みを実現している。先読みするブロックはアクセストレース情報を基に決定している。

【００１０】データの先読み（プリフェッチ）機能に関しては、以下のような技術が存在する。ＳＣＳＩ－２詳細解説（菅谷著，ＣＱ出版社，ＩＳＢＮ４－７８９８－３５２３－５）の１７７ページ（文献４）には、ＳＣＳＩ－２規格として定義されているＰｒｅ－Ｆｅｔｃｈコマンドについての説明が記述されている。このコマンドは、記憶装置に対して、指定された論理ブロックアドレスから指定されたデータ長分のデータを記憶媒体から読み出してデータキャッシュメモリに格納することを指示するコマンドである。

【００１１】米国特許５８８７１５１（文献５）においては、１つのプリフェッチコマンドにおいて、プリフェッチすべき複数のブロックをリストとして受け取ることが可能な記憶装置に関する技術が公開されている。

【００１２】論文“Ｉｎｆｏｒｍｅｄ　Ｐｒｅｆｅｔｃｈｉｎｇ　ａｎｄ　Ｃａｃｈｉｎｇ”（Ｒ．Ｈ．Ｐａｔｔｅｒｓｏｎ他著，Ｐｒｏｃ．　ｏｆ　ｔｈｅ　１５ｔｈＡＣＭ　Ｓｙｍｐ．　ｏｎ　Ｏｐｅｒａｔｉｎｇ　Ｓｙｓｔｅｍ　Ｐｒｉｎｃｉｐｌｅｓ，ｐｐ．７９－９

５，１９９５年１２月）（文献６）においては、アプリケーションが発行する今後アクセスを行うファイルとアクセス先領域に関するヒントを用いて、計算機のＯＳにおいて計算機上のファイルキャッシュにデータをプリフェッチする機能とその制御方法について論じている。特に、既存のＲＤＢＭＳを修正したプログラムを用いた評価も行われており、本技術のＲＤＢＭＳに対する有効性も示されている。

【００１３】

【発明が解決しようとする課題】従来の技術には以下のような問題が存在する。

【００１４】文献１に記載の技術はＤＢＭＳにおける技術であり、文献の実施例中では記憶装置に対して相対ブロックアドレスを用いてアクセスしている。現在の大容量な記憶装置においては、記憶装置内にキャッシュメモリや複数の物理記憶装置を保持し、それらを制御する記憶装置制御装置により１つの記憶装置として動作しているものが存在する。このような記憶装置においては、内部制御の最適化によりより高いアクセス性能を発揮させる余地が存在するが、文献１の技術でＤＢＭＳはこのような記憶装置の内部処理の最適化までは考慮していない。

【００１５】文献２に記載の技術においては、ＤＢＭＳの動作特性を考慮していない。そのため、ＤＢＭＳにより同時にアクセスされるデータを同じ物理記憶装置に配置してしまう可能性があり、これはＤＢＭＳのアクセス性能を低下させる要因となる。また、データの先読み等の記憶装置内キャッシュ制御の最適化については何も考慮していない。

【００１６】文献３に記載の記憶装置の高機能化によるＤＢＭＳの性能向上について論じている部分においては、記憶装置がアプリケーションレベルの知識のとしてのＲＤＢＭＳにおける問い合わせ処理の実行時の処理の実行プランを受け取る方法と、それを受け取った後にどのようなデータを用いてどのような処理を行う必要があるかの明確化が不十分である。

【００１７】文献３，文献４，文献５に記載の記憶装置外部からプリフェッチすべきブロックを教える技術に関しては、プリフェッチコマンドによるアクセスの実行優先順位が明確でない問題点が存在する。同一の物理記憶装置に記憶されているブロックに対する異なるプリフェッチコマンドが記憶装置に到着した場合、どちらを優先的に実行すべきかは状況により異なるが、ブロック指定のプリフェッチコマンドではどのブロックを優先すべきか明確でない。

【００１８】文献６に記載の技術に関しては、アプリケーションがヒントを発行する必要があり、既存のＤＢＭＳに対して適用する際にはプログラムの修正を要する。一般に、ＤＢＭＳに対しては極めて高い信頼性が要求される。ＤＢＭＳは複雑なプログラムであり修正は容易でないことやプログラムの修正は信頼性低下の要因であることを考えると、この技術は未対応な既存ＤＢＭＳに対して必ずしも適用できるものではない。

【００１９】本発明の第一の目的は、ＤＢＭＳが管理するデータを保持する記憶装置において、ＤＢＭＳ向けのアクセスの最適化を実施する記憶装置を実現することである。この記憶装置を用いることにより、既存のＤＢＭＳに対してプログラムの修正無しにＤＢＭＳ稼動システムの性能を向上させることができるようになる。

【００２０】本発明の第二の目的は、ＤＢＭＳが管理するデータを保持する記憶装置において、ＤＢのデータや処理に対する処理優先度を考慮したアクセスの最適化を実施する記憶装置を実現することである。ＤＢ毎や処理毎の処理優先度を考慮することにより、特定のＤＢに対する処理性能を保持するようなＤＢシステムを実現できるようになる。

【００２１】

【課題を解決するための手段】一般に、ＲＤＢＭＳは、与えられた問い合わせ処理（クエリ）を実施する前にその実行プランを作成し、それに従って処理を実施する。このクエリ実行プランは、どのデータをどのようにアクセスし、アクセス後にホスト上でどのような処理を実行するかの手順を示すものである。そこで、ＤＢＭＳの処理の実行情報である問い合わせ処理の実行プランを記憶装置自身が取得し、その情報と記憶装置内のデータ配置を考慮してこれからアクセスが行われるデータとそのアクセス順を予測し、それらのデータをあらかじめキャッシュメモリにプリフェッチしておくことにより、高いアクセス性能を有する記憶装置を実現する。

【００２２】ＤＢＭＳが直接、クエリ実行プランを記憶装置に対して与えることができない場合、同じ問い合わせ処理を実施する場合に作成されるクエリ実行プランは同じものとなることを利用する。多くのＲＤＢＭＳには与えられた問い合わせに対する実行プランを外部に出力する機能を有しており、問い合わせを実行する前にその機能によりクエリ実行プランを取得し、その後にＲＤＢＭＳに対して問い合わせを実行するようなプログラムを作成し、これを介して問い合わせ処理を実施するようにする。

【００２３】また、ＤＢのデータに対する処理優先度に関する情報を事前に取得しておき、プリフェッチ実行時にこれを考慮したプリフェッチやアクセス制御を実施する。特に、複数のＤＢに関するデータが同一の記憶装置上に存在する場合に、処理により高い優先度を要求するＤＢのデータに対して、より多くのキャッシュメモリとより高い物理記憶装置に対するアクセス割合を割り当て、高優先度ＤＢに対する処理性能の劣化を防ぐ。

【００２４】更に、クエリ実行プランを取得する際に、処理の優先度を指定するようにし、データに対する処理の優先度と同様に処理の優先度も考慮することにより高

い優先度を持つ処理の性能の劣化を防ぐ。

【００２５】

【発明の実施の形態】以下、本発明の実施の形態を説明する。なお、これにより本発明が限定されるものではない。

＜第一の実施の形態＞本実施の形態では、ＤＢＭＳが実行される計算機とデータキャッシュを保持する記憶装置が接続された計算機システムにおいて、記憶装置がＤＢＭＳに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報、ＤＢＭＳで実行されるクエリの実行プラン情報、ＤＢの処理優先順位情報を取得し、それらを用いて記憶装置がより好ましいアクセス性能を提供する。

【００２６】記憶装置は、ＤＢＭＳに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報、ＤＢＭＳで実行されるクエリの実行プランを利用することにより、ＤＢＭＳがこれからどのデータをどの順序でどのようにアクセスするかを把握することができる。そこで、この把握したアクセス方法に関する情報を利用して、あらかじめ利用される可能性が高いデータを記憶装置上のデータキャッシュ上に用意しておくことにより、ＤＢＭＳに対してより高いアクセス性能を提供する。また、ＤＢの処理優先順位情報を利用し、処理優先度が高いＤＢのデータや処理に対して、記憶装置が保持する物理記憶装置へアクセスを優先的に実施したり、また、データキャッシュの利用量をより多く割り当てたりすることにより、処理優先度が高いＤＢのデータや処理に対するアクセス性能を向上させる。

【００２７】図１は、本発明の第一の実施の形態における計算機システムの構成図である。本実施の形態における計算機システムは、ＤＢホスト８０ａ，８０ｂ、ＤＢクライアント８１、処理性能管理サーバ８２、記憶装置１０から構成される。ＤＢホスト８０ａ，８０ｂ、ＤＢクライアント８１、処理性能管理サーバ８２、記憶装置１０はそれぞれが保有するネットワークインターフェイス７８を通してネットワーク７９に接続されている。また、ＤＢホスト８０ａ，８０ｂ、記憶装置１０はそれぞれが保有するＩ／Ｏパスインターフェイス７０からＩ／Ｏパス７１を介してＩ／Ｏパススイッチ７２に接続され、これらを通して記憶装置１０とＤＢホスト８０ａ，８０ｂ間のデータ転送を行う。

【００２８】本実施の形態においては、記憶装置１０とＤＢホスト８０ａ，８０ｂ間のデータ転送を行うＩ／Ｏパス７１とネットワーク７９を異なるものとしているが、例えばｉＳＣＳＩのような計算機と記憶装置間のデータ転送をネットワーク上で実施する技術も開発されており、本実施の形態においてもこの技術を利用してもよい。このとき、記憶装置１０とＤＢホスト８０ａ，８０ｂにおいてＩ／Ｏパスインターフェイス７０が省かれ、計算機システム内からＩ／Ｏパス７１とＩ／Ｏパススイッチ７２が省かれる構成となる。

【００２９】記憶装置１０は、記憶領域を提供するもので、その記憶領域は記憶領域管理単位であるボリュームを用いて外部に提供し、ボリューム内の部分領域に対するアクセスや管理はブロックを単位として実行する。記憶装置１０は、ネットワークインターフェイス７８、Ｉ／Ｏパスインターフェイス７０、記憶装置制御装置１２、ディスクコントローラ１６、物理記憶装置１８から構成され、ネットワークインターフェイス７８、Ｉ／Ｏパスインターフェイス７０、記憶装置制御装置１２、ディスクコントローラ１６はそれぞれ内部バス２０により接続され、ディスクコントローラ１６と物理記憶装置１８は物理記憶装置バス２２により接続される。記憶装置制御装置１２は、ＣＰＵ２４とメモリ２６を有する。

【００３０】メモリ２６上には、記憶装置におけるキャッシュメモリとして利用するデータキャッシュ２８が割り当てられ、記憶装置を制御するためのプログラムである記憶装置制御プログラム５０が記憶される。また、メモリ２６上には、物理記憶装置１８の稼動情報である物理記憶装置稼動情報３２、データキャッシュ２８の管理情報であるデータキャッシュ管理情報３４、ＤＢＭＳ１１０ａ，１１０ｂが利用するＤＢデータに対する処理優先度とそれらを考慮したディスクＩ／Ｏの管理情報である処理優先度付ディスクＩ／Ｏ管理情報３６、ＤＢホスト８０ａ，８０ｂで実行されているＤＢＭＳ１１０ａ，１１０ｂで実行されるクエリの実行プランやそれを用いたプリフェッチの実行管理情報であるＤＢＭＳ実行情報３８、ＤＢホスト８０ａ，８０ｂで実行されているＤＢＭＳ１１０ａ，１１０ｂに関する情報であるＤＢＭＳデータ情報４０、記憶装置１０が提供するボリュームを物理的に記憶する物理記憶装置１８上の記憶位置の管理情報であるボリューム物理記憶位置管理情報４２を保持する。

【００３１】図中の記憶装置１０は、複数の物理記憶装置１８を有し、１つのボリュームに属するデータを複数の物理記憶装置１８に分散配置することが可能である。記憶装置制御プログラム５０は、ディスクコントローラ１６の制御を行うディスクコントローラ制御部５２、データキャッシュ２８の管理を行うキャッシュ制御部５４、記憶装置１０が提供するボリュームを物理的に記憶する物理記憶装置１８上の記憶位置の管理に関する処理を行う物理記憶位置管理部５６、Ｉ／Ｏパスインターフェイス７０の制御を行うＩ／Ｏパスインターフェイス制御部５８、ネットワークインターフェイス７８の制御を行うネットワークインターフェイス制御部６０を含む。

【００３２】ＤＢホスト８０ａ，８０ｂ、ＤＢクライアント８１、処理性能管理サーバ８２においては，それぞれＣＰＵ８４、ネットワークインターフェイス７８、メモリ８８を有し、メモリ８８上にオペレーティングシステム（ＯＳ）１００が記憶・実行されている。

【００３３】ＤＢホスト８０ａ，８０ｂはＩ／Ｏパスインターフェイス７０を有し、記憶装置１０が提供するボリュームに対してアクセスを実行する。ＯＳ１００内にファイルシステム１０４と１つ以上のボリュームからホストが利用する論理的なボリュームである論理ボリュームを作成するボリュームマネージャ１０２と、ファイルシステム１０４やボリュームマネージャ１０２により、ＯＳ１００によりアプリケーションに対して提供されるファイルや論理ローボリュームに記憶されたデータの記録位置等を管理するマッピング情報１０６を有する。ＯＳ１００が認識するボリュームやボリュームマネージャ１０２により提供される論理ボリュームに対して、アプリケーションがそれらのボリュームをファイルと等価なインターフェイスでアクセスするための機構であるローデバイス機構をＯＳ１００が有していても良い。

【００３４】図中の構成ではボリュームマネージャ１０２が存在しているが、本実施の形態においてはボリュームマネージャ１０２における論理ボリュームの構成を変更することはないので、ボリュームマネージャ１０２が存在せずにファイルシステムが記憶装置１０により提供されるボリュームを利用する構成に対しても本実施の形態を当てはめることができる。

【００３５】ＤＢホスト８０ａ，８０ｂのそれぞれのメモリ８８上ではＤＢＭＳ１１０ａ，１１０ｂが記憶・実行されている。ＤＢＭＳ１１０ａ，１１０ｂは内部にスキーマ情報１１４を有している。図中では、ＤＭＢＳ１１０ａ，１１０ｂが１台のホストに１つのみ動作しているが、後述するように、ＤＢＭＳ１１０ａ，１１０ｂ毎の識別子を用いて管理を行うため、１台のホストにＤＢＭＳが複数動作していても本実施の形態にあてはめることができる。

【００３６】ＤＢホスト８０ａ上ではＤＢＭＳ情報取得・通信プログラム１１８とクエリプラン取得プログラム１２０が動作している。一方、ＤＢホスト８０ｂ上ではＤＢＭＳ情報取得・通信プログラム１１８とクエリプラン取得プログラム１２０が提供する機能をＤＢＭＳ１１０ｂ中のＤＢＭＳ情報収集・通信部１１６が提供する。

【００３７】ＤＢクライアント８１のメモリ８８上では、ＤＢＭＳ１１０ａ，１１０ｂに対して処理要求を発行するＤＢＭＳフロントエンドプログラム１２６が記憶・実行される。図中では、ＤＢＭＳフロントエンドプログラム１２６はＤＢホスト８０ａ，８０ｂと異なる計算機上で動作しているが、ＤＢホスト８０ａ，８０ｂ上で動作していても本実施の形態に当てはめることができる。

【００３８】処理性能管理サーバ８２のメモリ８８上ではホスト情報設定プログラム１３０と処理性能管理プログラム１３２が記憶・実行される。図中では、ホスト情報設定プログラム１３０と処理性能管理プログラム１３２はＤＢホスト８０ａ，８０ｂ、ＤＢクライアント８１と異なる計算機上で動作しているが、それぞれ任意のＤＢホスト８０ａ，８０ｂ、ＤＢクライアント８１上で動作していても本実施の形態に当てはめることができる。

【００３９】図２はＤＢホスト８０ａ，８０ｂのＯＳ１００内に記憶されているマッピング情報１０６を示す。マッピング情報１０６中には、ボリュームローデバイス情報５２０、ファイル記憶位置管理情報５３０と論理ボリューム構成情報５４０が含まれる。ボリュームローデバイス情報５２０中にはＯＳ１００においてローデバイスを指定するための識別子であるローデバイスパス名５２１とそのローデバイスによりアクセスされる記憶装置１０が提供するボリュームあるいは論理ボリュームの識別子であるローデバイスボリューム名５２２の組が含まれる。

【００４０】ファイル記憶位置情報５３０中には、ＯＳ１００においてファイルを指定するための識別子であるファイルパス名５３１とそのファイル中のデータ位置を指定するブロック番号であるファイルブロック番号５３２とそれに対応するデータが記憶されている記憶装置１０が提供するボリュームもしくは論理ボリュームの識別子であるファイル配置ボリューム名５３３とそのボリューム上のデータ記憶位置であるファイル配置ボリュームブロック番号５３４の組が含まれる。

【００４１】論理ボリューム構成情報５４０中にはボリュームマネージャ１０２により提供される論理ボリュームの識別子である論理ボリューム名５４１とその論理ボリューム上のデータの位置を示す論理ボリュームブロック番号５４２とその論理ブロックが記憶されているボリュームの識別子であるボリューム名５０１とボリューム上の記憶位置であるボリュームブロック番号５１２の組が含まれる。マッピング情報１０６を取得するには、ＯＳ１００が提供している管理コマンドの実行や情報提供機構の利用、場合によっては参照可能な管理データの直接解析等を行う必要がある。

【００４２】図３はＤＢＭＳ１１０ａ，１１０ｂ内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報１１４を示す。スキーマ情報１１４には、表のデータ構造や制約条件等の定義情報を保持する表定義情報５５１、索引のデータ構造や対象である表等の定義情報を保持する索引定義情報５５２、利用するログに関する情報であるログ情報５５３、利用する一時表領域に関する情報である一時表領域情報５５４、管理しているデータのデータ記憶位置の管理情報であるデータ記憶位置情報５５５とデータをアクセスする際の並列度に関する情報である最大アクセス並列度情報５５７を含む。

【００４３】データ記憶位置情報５５５中には、表、索引、ログ、一時表等のデータ構造の識別子であるデータ構造名５６１とそのデータを記憶するファイルまたはローデバイスの識別子であるデータファイルパス名５６２と

その中の記憶位置であるファイルブロック番号５６３との組が含まれる。最大アクセス並列度情報５５７には、データ構造名５６１と、そのデータ構造にアクセスする際の一般的な場合の最大並列度に関する情報である最大アクセス並列度５６９の組が含まれる。スキーマ情報１１４を外部から取得するには、管理ビューとして外部に公開されているものをＳＱＬ等のデータ検索言語を用いて取得したり、または、専用の機構を用いて取得したりすることができる。

【００４４】図４は記憶装置１０内に保持されているボリューム物理記憶位置管理情報４２を示す。ボリューム物理記憶位置管理情報４２中には、ボリューム名５０１とそのボリューム上のデータ記憶位置であるボリューム論理ブロック番号５１２とその論理ブロックが記憶されている物理記憶装置１８の識別子である物理記憶装置名５０２と物理記憶装置１８上の記憶位置である物理ブロック番号５１４の組のデータが含まれる。

【００４５】図５に記憶装置１０内に保持されている物理記憶装置稼動情報３２を示す。物理記憶装置稼動情報３２中には、記憶装置１０が提供するボリュームの識別子であるボリューム名５０１とそのボリューム名５０１を持つボリュームのデータを保持する物理記憶装置１８の識別子である物理記憶装置名５０２、そしてボリューム名５０１を持つボリュームが物理記憶装置名５０２を持つ物理記憶装置１８に記憶しているデータをアクセスするための稼動時間のある時刻からの累積値である累積稼働時間５０３、稼動率５９４計算のために前回利用した累積稼働時間５０３の値である旧累積稼動時間５９３とある一定時間内の動作時間の割合を示す稼働率５９４の組と、稼働率５９４計算のために前回累積稼動時間を取得した時刻である前回累積稼動時間取得時刻５９５を含む。

【００４６】ディスクコントローラ制御部５２はディスクコントローラ１６を利用して物理記憶装置１８へのデータアクセスする際の開始時刻と終了時刻を取得し、そのアクセスデータがどのボリュームに対するものかを判断して開始時刻と終了時刻の差分を稼動時間として対応するボリューム名５０１と物理記憶装置名５０２を持つデータの組の累積稼働時間５０３に加算する。ディスクコントローラ制御部５２は一定間隔で以下の処理を行う。累積稼働時間５０３と旧累積稼働時間５９３、前回累積稼動時間取得時刻５９５と現データ取得時刻を用いて前回累積稼動時間取得時刻５９５と現データ取得時刻間の稼働率５９４を計算・記憶する。その後、取得した累積稼働時間５０３を旧累積稼働時間５９３に、現データ取得時刻を前回累積稼動時間取得時刻５９５に記憶する。

【００４７】図６に記憶装置１０内に保持されているＤＢＭＳデータ情報４０を示す。ＤＢＭＳデータ情報４０中には、ＤＢＭＳスキーマ情報７１１、データ構造物理記憶位置情報７１２を含む。

【００４８】ＤＢＭＳデータ情報４０中に含まれるデータは、ＤＢホスト８０ａ，８０ｂ上に存在するデータを利用する必要があるものが含まれる。記憶装置１０は記憶装置１０の外部に存在する情報を処理性能管理サーバ８２で動作するホスト情報設定プログラム１３０を利用して取得する。ホスト情報設定プログラム１３０はネットワーク７９を通し、ＤＢホスト８０ａ上で実行され、マッピング情報１０６等必要となる情報の収集処理を実施するＤＢＭＳ情報取得・通信プログラム１１８や、ＤＢホスト８０ｂ上で実行されているＤＢＭＳ１１０ｂ中のＤＢＭＳ情報取得・通信プログラム１１８と等価な機能を実現するＤＢＭＳ情報収集・通信部１１６を利用して必要な情報を収集する。

【００４９】ホスト情報設定プログラム１３０は情報取得後、必要ならば記憶装置１０に情報を設定するためのデータの加工を行い、ネットワーク７９を通して記憶装置１０に転送する。記憶装置１０においては、ネットワークインターフェイス制御部６０が必要な情報が送られてきたことを確認し、キャッシュ制御部５４に渡し、必要な加工を行った後にその情報をＤＢＭＳデータ情報４０中の適切な場所に記憶する。

【００５０】前述のように、ホスト情報設定プログラム１３０は任意のＤＢホスト８０ａ，８０ｂ上で実行されてもよい。あるいは、キャッシュ制御部５４がホスト情報設定プログラム１３０の情報収集機能を有してもよい。これらの場合は、ＤＢホスト８０ａ，８０ｂから情報を転送する際にＩ／Ｏパス７１を通して行ってもよい。この場合、特定の領域に対する書き込みが特定の意味を持つ特殊なボリュームを記憶装置１０はＤＢホスト８０ａ，８０ｂに提供し、そのボリュームに対する書き込みがあった場合にＩ／Ｏパスインターフェイス制御部は情報の転送があったと判断し、その情報をキャッシュ制御部５４に渡し、必要な加工を行った後にその情報をＤＢＭＳデータ情報４０中の適切な場所に記憶する。

【００５１】情報の収集処理に関しては、記憶装置１０が必要になったときに外部にデータ転送要求を出す方法と、データの変更があるたびに外部から記憶装置１０に変更されたデータを送る方法の２種類ともに利用することができる。ただし、ＤＢＭＳ１１０ａ，１１０ｂにおけるクエリ実行プランに関しては、実行する処理が明らかになった時点で受け取る必要があるので、記憶装置１０は処理性能管理プログラム１３２もしくはクエリプラン取得プログラム１２０もしくはＤＢＭＳ１１０ｂが与えるものを受動的に受け取る必要がある。

【００５２】図７にＤＢＭＳデータ情報４０中に含まれるＤＢＭＳスキーマ情報７１１を示す。ＤＢＭＳスキーマ情報７１１は、ＤＢＭＳデータ構造情報６２１、ＤＢＭＳデータ記憶位置情報６２２、ＤＢＭＳパーティション化表・索引情報６２３、ＤＢＭＳ索引定義情報６２

４、ＤＢＭＳホスト情報６２６、ＤＢＭＳホストマッピング情報６２７を含む。ＤＢＭＳデータ構造情報６２１はＤＢＭＳ１１０ａ，１１０ｂで定義されているデータ構造に関する情報で、ＤＢＭＳ１１０ａ，１１０ｂの識別子であるＤＢＭＳ名６３１、ＤＢＭＳ１１０ａ，１１０ｂ内の表・索引・ログ・一時表領域等のデータ構造の識別子であるデータ構造名５６１、データ構造の種別を表すデータ構造種別６４０、データ記憶位置情報から求めることができるデータ構造が利用する総データ量を示すデータ構造データ量６４１、そのデータ構造をアクセスする際の最大並列度に関する情報である最大アクセス並列度５６９の組を保持する。このとき、データ構造によっては最大アクセス並列度５６９の値を持たない。

【００５３】ＤＢＭＳデータ記憶位置情報６２２はＤＢＭＳ名６３１とそのＤＢＭＳにおけるデータ記憶位置管理情報５５５であるデータ記憶位置管理情報６３８の組を保持する。ＤＢＭＳパーティション化表・索引情報６２３は、１つの表や索引をある属性値により幾つかのグループに分割したデータ構造を管理する情報で、パーティション化されたデータ構造が所属するＤＢＭＳ１１０ａ，１１０ｂの識別子であるＤＢＭＳ名６３１と分割化される前のデータ構造の識別子であるパーティション元データ構造名６４３と分割後のデータ構造の識別子であるデータ構造名５６１とその分割条件を保持するパーティション化方法６４４の組を保持する。今後、パーティション化されたデータ構造に関しては、特に断らない限り単純にデータ構造と呼ぶ場合にはパーティション化後のものを指すものとする。

【００５４】ＤＢＭＳ索引定義情報６２４には、ＤＢＭＳ名６３１、索引の識別子である索引名６３５、その索引のデータ形式を示す索引タイプ６３６、その索引がどの表のどの属性に対するものかを示す対応表情報６３７組を保持する。ＤＢＭＳホスト情報６２６は、ＤＢＭＳ名６３１を持つＤＢＭＳ１１０ａ，１１０ｂがどのホスト上で実行されているかを管理するもので、ＤＢＭＳ名６３１とＤＢＭＳ実行ホストの識別子であるホスト名６５１の組を保持する。

【００５５】ＤＢＭＳホストマッピング情報６２７はＤＢホスト８０ａ，８０ｂのＯＳ１００内に記憶されているマッピング情報１０６収集したもので、ホスト名６５１とそのホストにおけるマッピング情報１０６を保持するマッピング情報６４８の組からなる。ＤＢＭＳホスト情報６２６はシステム構成情報で管理者が設定するものである。ＤＢＭＳスキーマ情報７１１中のその他のデータはＤＢＭＳ１１０ａ，１１０ｂが管理しているスキーマ情報１１４の中から必要な情報を取得して作成する。

【００５６】図８にＤＢＭＳデータ情報４０中に含まれるデータ構造物理記憶位置情報７１２を示す。データ構造物理記憶位置情報７１２はＤＢＭＳ１１０ａ，１１０ｂに含まれるデータ構造が記憶装置１０内でどの物理記憶装置１８のどの領域に記憶されるかを管理するもので、データ構造を特定するＤＢＭＳ名６３１とデータ構造名５６１、そのデータ構造内のブロックの通番であるデータ構造ブロック通番７１６、その外部からのアクセス領域を示すボリューム名５０１とボリュームブロック番号５１２、その物理記憶装置１８上の記憶位置を示す物理記憶装置名５０２と物理ブロック番号５１４の組を保持する。この情報は、ＤＢＭＳスキーマ情報７１１内のＤＢＭＳデータ記憶位置情報６２２とＤＢＭＳホストマッピング情報６２７とボリューム物理記憶位置メイン情報５１０を参照して、対応する部分を組み合わせることにより作成する。

【００５７】ＤＢＭＳ１１０ａ，１１０ｂ毎にシーケンシャルアクセスの方法が定まっている。ＤＢＭＳ名６３１とデータ構造名５６１により特定されるデータ構造毎に、シーケンシャルアクセス時のアクセス順を保持するようにソートしたデータをデータ構造物理記憶位置情報７１２は保持する。そして、データ構造毎のシーケンシャルアクセス順にしたがってデータ構造ブロック通番７１６を割り当てる。ここでは、対象とするＤＢＭＳ１１０ａ，１１０ｂの種類を絞り、あらかじめデータ構造物理記憶位置情報７１２を作成するプログラムがＤＢＭＳ１１０ａ，１１０ｂにおけるシーケンシャルアクセス方法を把握し、シーケンシャルアクセス順でソートされたデータを作成する。

【００５８】本実施の形態のＤＢＭＳ１１０ａ，１１０ｂにおけるシーケンシャルアクセス方法は以下の方法に従うものとする。あるデータ構造のデータをシーケンシャルアクセスする場合に、データ構造が記憶されているデータファイル名５６２とファイルブロック番号５６３を昇順にソートしその順序でアクセスを実行する。その他にシーケンシャルアクセス方法の決定方法としては、データファイルを管理する内部通番とファイルブロック番号５６３の組を昇順にソートした順番にアクセスする方法等が存在し、それらを利用したシーケンシャルアクセス方法の判断を実施してもよい。

【００５９】図９に記憶装置１０内に保持されているＤＢＭＳ実行情報３８を示す。ＤＢＭＳ実行情報３８中には、現在有効な実行情報を管理する実行情報ＩＤ管理情報８００と取得したクエリ実行プランをもとに作成するデータのアクセス方法を保持するＤＢＭＳ処理計画情報８０５、シーケンシャルアクセスに対するプリフェッチを実行する際の管理情報であるシーケンシャルプリフェッチ情報８１０、Ｂ－Ｔｒｅｅ索引を解釈することによるプリフェッチを実行する際の管理情報であるＢ－Ｔｒｅｅ索引プリフェッチ情報８２０、プリフェッチを実行する際のアクセスパターンを把握する際に利用するデータ構造アクセス情報８２５を含む。

【００６０】実行情報ＩＤ管理情報８００中には実行情報の識別子である実行情報ＩＤ８０１とその実行情報で

示される処理を実施するＤＢＭＳ１１０ａ，１１０ｂの識別子であるＤＢＭＳ名６３１、その実行情報を作成するのに用いたクエリプランの識別子であるクエリプランＩＤ８０２、そのクエリの処理優先度であるクエリ優先度８０３、その実行情報をもとに実行するプリフェッチ用のキャッシュの管理情報のＩＤであるＤＢＭＳキャッシュ管理ＩＤ８１１の組を有する。各エントリ値が有効値の場合にはそのエントリは有効でそうでない場合にはそのエントリは無効である。

【００６１】ＤＢＭＳ処理計画情報８０５中には、実行情報ＩＤ８０１と実行情報内の管理のための実行情報内部通番８０６とクエリ実行プランから識別されるデータ構造名５６１、クエリ実行プランから識別されるクエリ実行プラン中のデータ構造へのアクセス順序である実行順序８０７、そのデータ構造へのアクセス方法であるアクセスタイプ８０８の組を保持する。この情報に関しては、後述するクエリ実行プランに関する情報をもとに設定する。

【００６２】シーケンシャルプリフェッチ情報８１０中には，ＤＢＭＳ処理計画情報のエントリを特定する実行情報ＩＤ８０１と実行情報内部通番８０６、プリフェッチ用キャッシュの管理領域の識別子であるＤＢＭＳキャッシュ管理ＩＤ８１１、そのプリフェッチ管理情報によりプリフェッチが実行されるデータ構造の領域を示すデータ領域範囲８１２、どこまでプリフェッチを実行したかを管理するプリフェッチポインタ８１３、そのプリフェッチ領域へホストからの実アクセスの状況を示すアクセス状況８１４の組を含む。データ領域範囲８１２が示す領域は、データ構造ブロック通番７１６を用いて領域範囲を示す。アクセス状況８１４は、ホストからのアクセスが実行されていない“未実行”、ホストからのアクセスが存在した“実行”のいずれかを保持する。

【００６３】Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０中には、ＤＢＭＳ処理計画情報のエントリを特定する実行情報ＩＤ８０１と実行情報内部通番８０６、プリフェッチ用キャッシュの管理領域の識別子であるＤＢＭＳキャッシュ管理ＩＤ８１１，索引により選択するデータの条件を保持する索引アクセス条件８２１の組を保持する。

【００６４】データ構造アクセス情報８２５中には、データ構造を特定するＤＢＭＳ名６３１とデータ構造名５６１、データ構造毎の最も最近に行われたある一定数のアクセス履歴情報であるデータアクセス情報８２６を含む。データアクセス情報８２６はアクセス先としてデータ構造ブロック通番７１６を用いる。また、アクセス先とアクセスサイズは組にしてＦＩＦＯアルゴリズムで管理する。

【００６５】記憶装置１０は、データキャッシュをある一定サイズの領域であるセグメントと呼ぶ管理単位を用いて管理する。図１０に記憶装置１０内に保持されているデータキャッシュ管理情報３４を示す。データキャッシュ管理情報３４中には、データキャッシュ３４のセグメントの状態を示すキャッシュセグメント情報７２０とキャッシュセグメントの再利用対象選定に利用するキャッシュセグメント利用管理情報７４０と、ＤＢＭＳ実行情報３８を利用したプリフェッチ実行のために割り当てられたキャッシュセグメントを管理するＤＢＭＳデータキャッシュ管理情報８３０を含む。

【００６６】キャッシュセグメント情報７２０中には、セグメントの識別子であるセグメントＩＤ７２１と、そのセグメントに記憶されているデータ領域を示すボリューム名５０１とボリューム論理ブロック番号５１２、そして、セグメントの状態を示すステータス情報７２２、後述するセグメントの管理に利用するリストの情報であるリスト情報７２３を含む。

【００６７】ステータス情報７２２が示すセグメントの状態としては、物理記憶装置１８上にセグメント内のデータと同じデータが記憶されている“ノーマル”、セグメント内にのみ最新のデータが存在する“ダーティ”、セグメント内のデータに対する書き込み要求発行中を示す“ライト”、セグメント内に有効なデータが存在しない“インバリッド”が存在する。リスト情報７２３には、現在そのセグメントが属するリストの識別子と、そのリストのリンク情報が記憶される。図中では、リストは双方向リンクリストであるとしている。

【００６８】キャッシュセグメント利用管理情報７４０中には、キャッシュセグメントの再利用対象選定に利用する２種類の管理リストであるメインＬＲＵリスト、再利用ＬＲＵリストの管理情報として、メインＬＲＵリスト情報７４１、再利用ＬＲＵリスト情報７４３が、現在のキャッシュセグメント情報７２０中のステータス情報７２２がダーティに設定されている数を示すダーティセグメント数カウンタ７４６とＤＢＭＳ実行情報３８を利用したプリフェッチを実行のために割り当てられたキャッシュセグメント数を示すＤＢＭＳプリフェッチ用割り当てセグメント数７４７が記憶される。メインＬＲＵリスト情報７４１、再利用ＬＲＵリスト情報７４３は、それぞれリストの先頭であるＭＲＵセグメントＩＤ、最後尾であるＬＲＵセグメントＩＤ、そのリストに属するセグメント数を記憶する。

【００６９】ＤＢＭＳデータキャッシュ管理情報８３０中には、エントリの識別子であるＤＢＭＳキャッシュ管理ＩＤ８１１とこのエントリにより管理されるリストの先頭セグメント、最後尾セグメント、セグメント数を表す先頭セグメントＩＤ８３１、最後尾セグメントＩＤ８３２、領域セグメント数８３３の組を保持する。領域セグメント数８３３の値でそのエントリが利用中かどうかを判断し、その値が０以上の場合には利用中と判断し、負の値の場合には未利用と判断する。

【００７０】図１１に記憶装置１０内に保持されている

処理優先度付ディスクＩ／Ｏ管理情報３６を示す。処理優先度付ディスクＩ／Ｏ管理情報３６はデータ構造の処理優先度に関する情報であるＤＢＭＳデータ構造処理優先度情報８４０とデータ構造に対して割り当てられた処理優先度をもとにした制御を行う際に利用する処理情報設定情報８５０、物理記憶装置１８へのデータアクセスを実行する際にディスクコントローラ１６に対するＩ／Ｏ要求の発行管理に利用するディスクＩ／Ｏ実行管理情報８６０を含む。

【００７１】ＤＢＭＳデータ構造処理優先度情報８４０中には、データ構造を特定するＤＢＭＳ名６３１とデータ構造名５６１、そのデータ構造に与えられたデフォルトの処理優先度である処理優先度８４１、ＤＢＭＳ実行情報３８で管理される実行情報中のクエリ優先度８０３を加味したデータ構造の処理優先度である実効処理優先度８４２の組と、データ構造に属さないその他の一般データに対するアクセスに与えられる処理優先度である一般データ処理優先度８４５を保持する。

【００７２】処理情報設定情報８５０中には、処理優先度８４１とその処理優先度に割り当てられる物理記憶装置１８へのＩ／Ｏ数の比率である割り当てＩ／Ｏ比率８５１、その処理優先度を持つ領域におけるクエリ毎にプリフェッチ用に割り当てるキャッシュセグメント数を示す領域プリフェッチ割り当て量８５２、その処理優先度を持つと判断されたクエリに対するプリフェッチ用に割り当てるキャッシュセグメント数の最大値を示すクエリプリフェッチ最大量８５３の組を保持する。

【００７３】ディスクＩ／Ｏ実行管理情報８６０中は、物理記憶装置１８の識別子である物理記憶装置５０２とそこへのディスクＩ／Ｏ発行を管理するために利用するＩ／Ｏ実行管理情報８６１の組を含む。Ｉ／Ｏ実行管理情報８６１中は、読出しアクセス要求を溜めるキューの配列である読出しキュー配列８６３と書込みアクセス要求を溜めるキューである書込みキュー８６４、処理優先度毎のＩ／Ｏ数割り当てを管理するために用いる処理優先度別処理残数カウンタ８６６を含む。

【００７４】読出しキュー配列８６３へのアクセスは、処理優先度８３１とホストアクセス要求に答えるためのものかプリフェッチ実施のためのものかを示す指示子８６７によりアクセスすべきアクセスキューを指定する。処理優先度別処理残数カウンタ８６６中には、処理優先度８３１と対応する処理優先度におけるアクセス可能残数を示す処理残りカウント値８６８の組を有する。

【００７５】続いて、クエリの実行プラン利用した記憶装置１０における制御方式を説明する。ＤＢＭＳ１１０ａ，１１０ｂは、処理を与えた場合にその結果を得るために内部的にどのような処理がどのような順番で実行されるかを示すクエリ実行プランを外部に提供する機能を有する。この機能により取得したクエリ実行プランを記憶装置１０に与えることにより、記憶装置１０はこれからどのような処理がＤＢＭＳ１１０ａ，１１０ｂで行われるかを把握することができる。その情報を用いてデータをあらかじめデータキャッシュ２８にプリフェッチしたり、あるいは、処理実行中にもうアクセスされないデータキャッシュ２８上のデータを把握してそれらのデータを保持しているキャッシュセグメントを優先的に再利用することにより、より高いアクセス性能を得る。

【００７６】クエリ実行プランの例として図１２中にクエリ８７１とその処理を実現するためにＤＢＭＳ１１０ａが作成したクエリ実行プラン８７２を図示したものを示す。図に示されるように、クエリ実行プラン８７２はクエリ８７１の結果を得るために内部で実行する細分化された処理をノードとする木構造で表現することが可能である。図中でデータは末端から幹の方向に向かって流れていく。処理ノード８７５ａ，８７５ｂ，８７５ｃ，８７５ｄ，８７５ｅ，８７５ｆ，８７５ｇ，８７５ｈはクエリ８６１で実行される細分化された処理を表し、枝８７６は処理間のデータの流れ関係を示す。処理グループ８７７ａ，８７７ｂ，８７７ｃはＤＢＭＳ１１０ａにおいて同時に実行される可能性がある処理の組を表し、１つの処理グループに属する処理が完了するまで他の処理グループに属する処理が実行されることはない。

【００７７】処理グループ内での処理は、処理グループ内で行われる処理の内容とそれらの処理に利用されるデータの流れによりその実行順序は定まる。クエリ実行計画８７２では、まず処理グループ８７７ｂにより、表Ｔ３が全走査される。続いて処理グループ８７７ｃにより、表Ｔ４の全走査が行われその結果をもとに属性値Ｍの値が１００未満の組が選択される。この表の全走査と属性値Ｍによるデータの選択は並行して実行される。続いて、処理グループ８７７ａの処理が行われる。処理グループ８７７ｂと処理グループ８７７ｃの処理結果のハッシュ結合処理を実施する。その結果を用いて索引Ｉｎｄ１－１を参照して表Ｔ１内の対応データを検索するネストループ結合を実施し、その結果から表Ｔ１の属性Ｂの値の総和を求める。処理グループ８７７ａ内の処理は並行して実施される。

【００７８】図１３はクエリ実行プラン８７２が作成されたときに、記憶装置１０に与えられるクエリ実行プランに関する情報であるクエリプラン情報８８０を示す。クエリプラン情報８８０中には、そのクエリ実行プランを持つ処理が行われるＤＢＭＳ１１０ａ，１１０ｂの識別子であるＤＢＭＳ名６３１とそのクエリ実行プランの識別子であるクエリプランＩＤ８０２、そのクエリ実行プランを持つ処理に与えられた処理の優先度であるクエリ優先度８０３、クエリの実行プランに関する詳細情報を保持するクエリ実行プラン情報８８１が含まれる。

【００７９】クエリ実行プラン情報８８１には、クエリ実行プラン８７２を木構造で表現した場合の処理ノード８７５の識別子８８３とその親の処理ノード８７５の識

別子８８４、ハッシュ結合・ネステッドループ結合・ソートマージ結合・全表走査・表アクセス・索引アクセス・フィルタ・ソート・総和演算等のその処理ノード８７５で実施されるノード処理内容８８５、処理ノード８７５でデータ構造に対するアクセスを実施する処理の場合のアクセス先のデータ構造名５６１であるアクセスデータ構造８８６、処理ノード８７５が所属する処理グループ８７７の処理グループ間の処理の実行順序を示す処理順序８８７、そして、結合演算における結合処理条件や索引アクセスにおけるデータ検索条件，データの並列アクセス時のデータ分割方法等の処理ノード８７５で実行される処理の詳細情報であるノード処理詳細８８８を含む。なお、アクセスデータ構造８８６はノード処理内容８８５がデータに対するアクセスではないエントリには無効値を入れる。ノード処理詳細８８８は必ずしも含まれなくともよい。

【００８０】クエリプラン情報８８０を記憶装置１０が受け取るときに、ＤＢクライアント８１上のＤＢＭＳフロントエンドプログラム１２６がＤＢＭＳ１１０ａに処理要求を出した場合の処理手順は以下のようになる。まず、ＤＢＭＳフロントエンドプログラム１２６はネットワーク７９を通してクエリプラン取得プログラム１２０に対してＤＢＭＳ１１０ａ上での処理の実行を依頼する。このとき、クエリ優先度８０３も指定する。また、クエリプラン取得プログラム１２０は処理を依頼されたときに、ＤＢＭＳフロントエンドプログラム１２６がどのようなものかの識別情報を作成する。

【００８１】クエリプラン取得プログラム１２０は、ＤＢＭＳフロントエンドプログラム１２６が依頼した処理に対するクエリ実行プラン８７２をＤＢＭＳ１１０ａから取得し、指定されたクエリ優先度８０３と処理を実行するＤＢＭＳ１１０ａのＤＢＭＳ名６３１、ＤＢＭＳフロントエンドプログラム１２６の識別情報、そして取得したクエリ実行プラン８７２とクエリプラン取得プログラム１２０が付加するクエリ実行プラン８７２の識別子であるクエリプランＩＤ８０２をネットワーク７９を介して処理性能管理サーバ８２上の処理性能管理プログラム１３２に送る。クエリ実行プラン８７２その他情報を受け取った処理性能管理プログラム１３２は、取得したクエリ優先度８０３とＤＢＭＳフロントエンドプログラム１２６の識別情報、その他処理優先度の設定をもとに記憶装置１０に与えるクエリ優先度８０３を決定し、その他取得した情報を用いてクエリプラン情報８８０を作成し、ネットワーク７９を通して記憶装置１０に送信する。

【００８２】クエリプラン取得プログラム１２０は、クエリ実行プラン８７２その他情報を処理性能管理プログラム１３２に送った後に、ＤＢＭＳフロントエンドプログラム１２６が依頼した処理をＤＢＭＳ１１０ａに対して送る。その結果をクエリプラン取得プログラム１２０が取得し、ＤＢＭＳフロントエンドプログラム１２６に結果を返す。

【００８３】その後、ＤＢＭＳ名６３１とクエリプランＩＤ８０２で識別されるクエリ実行プラン８７２により実行された処理が完了したことをネットワーク７９を介して処理性能管理プログラム１３２に伝える。クエリ実行プラン８７２に対応する処理の完了報告を受けた処理性能管理プログラム１３２はＤＢＭＳ名６３１とクエリプランＩＤ８０２で識別されるクエリプラン情報８８０に対応する処理が完了したことをネットワーク７９を通して記憶装置１０に伝える。

【００８４】上記の例では、クエリプラン取得プログラム１２０はＤＢＭＳ１１０ａが動作するＤＢホスト８０ａ上で動作しているが、任意の計算機、つまり、任意のＤＢホスト８０ａ，８０ｂ、ＤＢクライアント８１上で動作してもよい。また，ＤＢＭＳフロントエンドプログラム１２６がクエリプラン取得プログラム１２０の機能を含んでいてもよい。

【００８５】ＤＢＭＳ１１０ｂにおいては、クエリプラン取得プログラム１２０の役割をＤＢＭＳ１１０ｂのＤＢＭＳ情報通信部１０６がその役割を果たす。ＤＢクライアント８１上のフロントエンドプログラム１２６がＤＢＭＳ１１０ｂに処理要求を出した場合の処理手順は以下のようになる。まず、ＤＢＭＳフロントエンドプログラム１２６はネットワーク７９を通してＤＢＭＳ１１０ｂへ処理の実行を依頼する。このとき、クエリ優先度８０３も指定する。また、ＤＢＭＳ１１０ｂは処理を依頼されたときに、ＤＢＭＳフロントエンドプログラム１２６の識別情報を作成する。

【００８６】ＤＢＭＳ１１０ｂはＤＢＭＳフロントエンドプログラム１２６が依頼した処理に対するクエリ実行プラン８７２を作成し、指定されたクエリ優先度８０３とＤＢＭＳ１１０ｂのＤＢＭＳ名６３１、ＤＢＭＳフロントエンドプログラム１２６の識別情報、そして作成したクエリ実行プラン８７２とＤＢＭＳ１１０ｂが付加するクエリ実行プラン８７２の識別子であるクエリプランＩＤ８０２をＤＢＭＳ情報通信部１０６を利用してネットワーク７９を介して処理性能管理サーバ８２上の処理性能管理プログラム１３２に送る。

【００８７】クエリ実行プラン８７２その他情報を受け取った処理性能管理プログラム１３０は、取得したクエリ優先度８０３とＤＢＭＳフロントエンドプログラム１２６の識別情報，その他処理優先度の設定をもとに記憶装置１０に与えるクエリ優先度８０３を決定し、その他取得した情報を用いてクエリプラン情報８８０を作成し、ネットワーク７９を通して記憶装置１０に送信する。

【００８８】ＤＢＭＳ１１０ｂは処理性能管理プログラム１３２に対して情報を送った後に依頼された処理を実行し、その結果をＤＢＭＳフロントエンドプログラム１

２６に返す。その後、ＤＢＭＳ名６３１とクエリプランＩＤ８０２で識別されるクエリ実行プラン８７２により実行された処理が完了したことをＤＢＭＳ情報通信部１０６を利用してネットワーク７９を介して処理性能管理プログラム１３２に伝える。クエリ実行プラン８７２に対応する処理の完了報告を受けた処理性能管理プログラム１３２はＤＢＭＳ名６３１とクエリプランＩＤ８０２で識別されるクエリプラン情報８８０に対応する処理が完了したことをネットワーク７９を通して記憶装置１０に伝える。

【００８９】前述のように、処理性能管理プログラム１３２は計算機システム内の任意の計算機、つまり、任意のＤＢホスト８０ａ，８０ｂ、ＤＢクライアント８１上で動作してもよい。これまで述べてきた方法では、クエリプラン取得プログラム１２０またはＤＢＭＳ１１０ｂは一旦処理性能管理プログラム１３２へ情報を送り、そこでクエリプラン情報８８０を作成し記憶装置１０に与えている。

【００９０】その代わりに、クエリプラン取得プログラム１２０またはＤＢＭＳ１１０ｂが直接クエリプラン情報８８０を作成し、それを記憶装置１０へ送ってもよい。このとき、ＤＢホスト８０ａ，８０ｂから情報を転送する際にＩ／Ｏパス７１を通して行ってもよい。この場合、特定の領域に対する書き込みが特定の意味を持つ特殊なボリュームを記憶装置１０はＤＢホスト８０ａ，８０ｂに提供し、そのボリュームに対する書き込みがあった場合にＩ／Ｏパスインターフェイス制御部は情報の転送があったと判断する。

【００９１】上記の方法では、記憶装置１０が受け取る情報はクエリプラン情報８８０中のクエリの実行プランに関する情報はクエリ実行プラン情報８８１であるが、処理性能管理プログラム１３２でＤＢＭＳ実行情報３８に設定するＤＢＭＳ処理計画情報８０５、シーケンシャルプリフェッチ情報８１０、Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０の内容を作成し送ってもよい。このとき、実行情報ＩＤ８０１、ＤＢＭＳキャッシュ管理ＩＤ８１１、プリフェッチポインタ８１３、アクセス状況８１４に入れる値は任意のものでよい。また、これらの情報の作成方法は記憶装置１０内で設定する場合のものと同じ方法を用い、それらについては後述する。

【００９２】記憶装置１０がネットワーク７９通してあるいはＩ／Ｏパス７１を通してクエリプラン情報８８０を受け取った際の処理を説明する。図１４にクエリプラン情報８８０を受け取った際の処理フローを示す。これは、ネットワークインターフェイス７８がクエリプラン情報８８０を受け取った場合にはネットワークインターフェイス制御部６０が、Ｉ／Ｏパスインターフェイス７０がクエリプラン情報８８０を受け取った場合にはＩ／Ｏパスインターフェイス制御部５８がキャッシュ制御部５４にクエリプラン情報８８０を受け取ったことを伝えることによりキャッシュ制御部５４が処理を開始する。ステップ２００１で処理を開始する。

【００９３】ステップ２００２では取得したクエリプラン情報８８０をもとにプリフェッチ可能領域を把握し、ＤＢＭＳ実行情報３８中に必要な情報を設定する。まず、実行情報ＩＤ管理情報８００に空きエントリを探して実行情報を保存するための実行情報ＩＤ８０１を取得し、そこにクエリプラン情報８８０をもとにＤＢＭＳ名６３１、クエリプランＩＤ８０２、クエリ優先度８０３を設定し、ＤＢＭＳデータキャッシュ管理情報８３０中の空きエントリを捜し、そのＤＢＭＳキャッシュ管理ＩＤ８１１を設定する。

【００９４】そして、ＤＢＭＳデータキャッシュ管理情報８３０中の選択したＤＢＭＳキャッシュ管理ＩＤを持つエントリ中の領域セグメント数を０に設定し、そのエントリを利用中に設定する。なお、これ以降のクエリプラン情報８８０を取得した際の処理においては、ＤＢＭＳ名６３１はクエリプラン情報８８０の値とし、また、ここで取得した実行情報ＩＤ８０１とＤＢＭＳ名６３１は本処理中で常に参照可能な状態にしておく。

【００９５】続いて、クエリ実行プラン情報８８１をもとにＤＢＭＳ処理計画情報８０５を設定する。この情報には、クエリ実行プラン情報８８１中でアクセスデータ構造８８６に有効値が入っているエントリを選択し、それらを処理順序８８７によって降順にソートする。更に、処理順序８８７に同じ値が入っているものに関しては、索引をアクセスした後に表のデータを読む等、プランノード名８８３とプラン親ノード名８８４、ノード処理内容８８５によってデータ間の依存関係を調べ、先にアクセスされるものを先に持ってくるようにソートする。

【００９６】これらのソート結果にしたがって昇順に実行情報内部通番８０６が割り当てられるようにクエリ実行プラン情報８８１のエントリの内容をＤＢＭＳ処理計画情報８０５に設定していく。データ構造名５６１にアクセスデータ構造８８６を、実行順序８０７には処理順序８８７の値を設定する。

【００９７】アクセスタイプ８０８は以下のように設定する。まず、ＤＢＭＳ名６３１とアクセスデータ構造８８６からクエリ実行プラン情報８８１のエントリに対応するデータ構造の内容を把握する。データ構造が木構造索引であるものには、それに対応するアクセスタイプ８０８には“木構造索引”を、データ構造が表の場合には、そのクエリ実行プラン情報８８１中のデータ依存関係を調べ、木構造索引をアクセスした結果をもとにアクセスする場合には“索引参照”を、その他の場合には“シーケンシャル”を設定する。

【００９８】次に、シーケンシャルプリフェッチ情報８１０を設定する。シーケンシャルプリフェッチ情報８１０は、直前に設定されたＤＢＭＳ処理計画情報８０５の

エントリのうち、アクセスタイプ８０８に“シーケンシャル”と設定されたエントリに対して設定される。まず、アクセスタイプ８０８に“シーケンシャル”と設定されたエントリを取り出す。ＤＢＭＳ処理計画情報８０５中の確認中のエントリからデータ構造名５６１を取り出す。ＤＢＭＳ名６３１とデータ構造名５６１を索引名６３５とみなしてＤＢＭＳ索引定義情報６２４を参照してデータ構造がビットマップ索引かどうかを確認し、そうである場合にはシーケンシャルプリフェッチ情報８１０に何も設定しない。

【００９９】ＤＢＭＳ１１０ａ，１１０ｂにおいては、最大アクセス並列度５６９によりシーケンシャル順にデータ構造を等分し、それらを並列にアクセスするもとのする。ＤＢＭＳ名６３１とデータ構造名６６１を用いてＤＢＭＳデータ構造情報６２１を参照してそのデータ構造における最大アクセス並列度５６９を求める。シーケンシャルプリフェッチ情報８１０中に、現在設定中の実行情報ＩＤ８０１と実行情報内部通番８０６を持つエントリを取得した最大アクセス並列度５６９の値の分だけ作成する。ＤＢＭＳ名６３１とデータ構造名５６１を用いてデータ構造物理記憶位置情報７１２を参照し、データ構造を記憶する全領域のデータ構造ブロック通番７１６を取得し、それを取得した最大アクセス並列度５６９の値で等分し、現在作成中のシーケンシャルプリフェッチ情報８１０のエントリのデータ領域範囲８１２にそれぞれ設定する。

【０１００】そして、それぞれのエントリにおいて、プリフェッチポインタ８１３にデータ領域範囲８１２に設定された値の先頭部分を設定し、アクセス状況８１４には“未実行”を入れる。更に、ＤＢＭＳデータキャッシュ管理情報８３０中の空きエントリを捜し、そのＤＢＭＳキャッシュ管理ＩＤ８１１を設定する。そして、ＤＢＭＳデータキャッシュ管理情報８３０中の選択したＤＢＭＳキャッシュ管理ＩＤ８１１を持つエントリの領域セグメント数８３３を０に設定する。

【０１０１】最後に、Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０を設定する。Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０は、直前に設定されたＤＢＭＳ処理計画情報８０５のエントリのうち、アクセスするデータ構造が木構造のものに対して作成する。まず、アクセスタイプ８０８に“木構造索引”と設定されたエントリを取り出す。そのエントリの実行情報内部通番８０６を取り出し、先に対応が決定されたクエリ実行プラン情報８８１中のエントリを特定し、ノード処理詳細８８８の内容を確認する。

【０１０２】その内容が木構造索引をアクセスする際の条件を含まないときには、Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０には何も設定しない。ノード処理内容詳細８８８の内容が木構造索引をアクセスする際の条件を含むとき、その条件が前の処理の結果に依存する場合にはＢ－Ｔｒｅｅ索引プリフェッチ情報８２０に何も設定しない。依存しない場合には、Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０にエントリを作成し、現在確認中のＤＢＭＳ処理計画情報８０５のエントリの実行情報ＩＤ８０１と実行情報内部通番８０５を値を作成したエントリに設定する。

【０１０３】そして、木構造索引をアクセスする際の条件を索引アクセス条件８２１に設定する。そして、ＤＢＭＳデータキャッシュ管理情報８３０中の空きエントリを捜し、そのＤＢＭＳキャッシュ管理ＩＤ８１１を設定する。そして、ＤＢＭＳデータキャッシュ管理情報８３０中の選択したＤＢＭＳキャッシュ管理ＩＤ８１１を持つエントリの領域セグメント数８３３を０に設定する。

【０１０４】ステップ２００３では、プリフェッチ処理に利用するキャッシュセグメント数の計算とそれに付随する処理を行う。まず現在設定中の実行情報ＩＤ８０１により実行情報ＩＤ管理情報８００を参照してクエリ優先度８０３を求める。続いて、現在設定中の実行情報ＩＤ８０１によりＤＢＭＳ処理計画情報８０５を参照しアクセスが実行されるデータ構造名５６１の組を求める。求めたデータ構造名５６１すべてに対して、ＤＢＭＳデータ構造処理優先度情報８４０を参照して処理優先度情報８４１を求める。

【０１０５】この値と前に取得したクエリ優先度８０３をもとにあらかじめ定められたルールを用いて各データ構造のこの処理における処理優先度を求める。そして、ＤＢＭＳデータ構造処理優先度情報８４０中の実効処理優先度情報８４２を参照して求めた各データ構造のこの処理における処理優先度と比較して、この処理における処理優先度の方が高い場合には実効処理優先度情報８４２をこの処理における処理優先度の値に更新する。

【０１０６】実行情報ＩＤ８０１を持つシーケンシャルプリフェッチ情報８１０の全エントリを検索し、実行情報ＩＤ８０１と選択された各エントリにおける実行情報内部通番８０６からアクセスが実行されるＤＢＭＳ処理計画情報８０５を参照してデータ構造名５６１を求め、このデータ構造に対応する先に求めたこの処理におけるデータ構造の処理優先度により処理優先度設定情報８５０から領域プリフェッチ割当量８５２を求め、そのシーケンシャルプリフェッチ情報８１０のエントリに対するプリフェッチ割り当て量とする。

【０１０７】続いて、実行情報ＩＤ８０１を持つＢ－Ｔｒｅｅ索引プリフェッチ情報８２０の全エントリを検索し、実行情報ＩＤ８０１と選択された各エントリにおける実行情報内部通番８０６からアクセスが実行されるＤＢＭＳ処理計画情報８０５を参照してデータ構造名５６１を求め、このデータ構造に対応する先に求めたこの処理におけるデータ構造の処理優先度により処理優先度設定情報８５０から領域プリフェッチ割り当て量８５２を求め、そのＢ－Ｔｒｅｅ索引プリフェッチ情報８２０の

エントリに対するプリフェッチ割り当て量とする。

【０１０８】実行情報ＩＤ８０１を持つシーケンシャルプリフェッチ情報８１０とＢ－Ｔｒｅｅ索引プリフェッチ情報８２０の全エントリに対するプリフェッチ割り当て量の総和を求める。先に求めたこの処理におけるデータ構造の処理優先度のうちの最も優先度が高い値を求め、これを用いて処理優先度設定情報８５０を参照して求めるクエリプリフェッチ最大量８５３とプリフェッチ割当量の総和を比較し、値の小さな方をこのクエリにおけるプリフェッチ用キャッシュ容量の希望値とする。

【０１０９】ステップ２００４では、先に求めた希望値の分だけプリフェッチ用キャッシュを割り当て可能か確認する。ＤＢＭＳプリフェッチ用割り当てセグメント数７４７と先に求めた希望値の和があらかじめ定めた閾値未満か確認し、閾値未満の場合には割り当て可能と判断してステップ２００５に進み、閾値以上の場合には割り当て不可能としてステップ２００６に進む。

【０１１０】ステップ２００５では、先に求めた希望値分のキャッシュセグメントを再利用ＬＲＵリストのＬＲＵ（Ｌｅａｓｔ　Ｒｅｃｅｎｔｌｙ　Ｕｓｅｄ：最も昔に使われた）側から選択してそこから外し、それにあわせてキャッシュセグメント情報７２０と再利用ＬＲＵリスト情報７４３を更新する。選択したキャッシュセグメントから新たなリストを作成し、それらを実行情報ＩＤ管理情報８００を参照して実行情報ＩＤ８０１に対応するＤＢＭＳキャッシュ管理ＩＤ８１１を求め、先に作成したリストの先頭セグメントＩＤ８３１と最後尾セグメントＩＤ８３２と領域セグメント数８３３を対応するエントリに記憶し、キャッシュセグメント情報７２０を更新し、これをプリフェッチに利用するキャッシュセグメントのプールとする。次にステップ２００８に進む。

【０１１１】ステップ２００６では、確保可能なキャッシュセグメントを再利用ＬＲＵリストから選択してそこから外し、それにあわせて再利用ＬＲＵリスト情報７４３とキャッシュセグメント情報７２０を更新する。選択したキャッシュセグメントから新たなリストを作成し、それらを実行情報ＩＤ管理情報８００を参照して実行情報ＩＤ８０１に対応するＤＢＭＳキャッシュ管理ＩＤ８１１を求め、それに対応するエントリに先に作成したリストの先頭セグメントＩＤ８３１と最後尾セグメントＩＤ８３２と領域セグメント数８３３を記憶し、キャッシュセグメント情報７２０を更新し、これをプリフェッチに利用するキャッシュセグメントのプールとする。

【０１１２】ステップ２００７では、クエリプラン情報８８０中のクエリ優先度８０３を求め、これより優先度が低い実行情報が存在するか実行情報ＩＤ管理情報８００を参照して確認し、存在する場合にはそれらが確保しているキャッシュセグメントを現在設定中の実行情報用にまわす。現在設定中の実行情報よりクエリ優先度が低いものをクエリ優先度が低い順に確認し、実行情報ＩＤ管理情報８００中の対応するＤＢＭＳキャッシュ管理ＩＤ８１１により識別されるＤＢＭＳデータキャッシュ管理情報８３０のエントリにより管理されるプリフェッチ用キャッシュセグメントのプールを確認し、そこにセグメントが存在する場合にはそれらを現在確保しているプリフェッチ用キャッシュのプールに回し、それにあわせてＤＢＭＳデータキャッシュ管理情報８３０を更新する。

【０１１３】これを先に求めた希望値分のキャッシュセグメントが現在設定中の実行情報に対応するプリフェッチ用キャッシュのプールに貯まるまで、あるいは、クエリプラン情報８８０中のクエリ優先度８０３より優先度が低い実行情報中のプリフェッチ用キャッシュのプールに存在するものすべて収集したらこのステップの処理を完了し、ステップ２００８に進む。このとき、現在設定中の実行情報のクエリ優先度８０３より優先度が低い実行情報中のプリフェッチ用キャッシュのプールに存在するものを集めても先に求めた希望値分まで貯まらない場合、更に、よりクエリ優先度が低い実行情報により既にプリフェッチが実行されたデータを保持するキャッシュセグメントを現在設定中の実行情報に対応するプリフェッチ用キャッシュのプールに持ち込む処理を追加してもよい。

【０１１４】ステップ２００８では、現在設定中の実行情報ＩＤ８０１を持つエントリのうち実行順序８０７の値が最も早いエントリより示されるデータ構造に対するプリフェッチ用キャッシュを割り当てる。現在設定中の実行情報ＩＤ８０１を持つエントリのうち実行順序８０７の値が最も早いエントリの実行情報内部通番８０６を求める。対応する実行情報内部通番８０６は複数存在する可能性がある。実行情報ＩＤ８０１により実行情報ＩＤ管理情報８００を参照して対応するＤＢＭＳキャッシュ管理ＩＤ８１１を求めてプリフェッチ用プールの情報を取得する。実行情報ＩＤ８０１と実行情報内部通番８０６に対応するシーケンシャルプリフェッチ情報８１０中のエントリに対しては、プールから最初に実行するプリフェッチ量だけキャッシュセグメントを割り当て、シーケンシャルプリフェッチ情報８１０からそのエントリにおけるＤＢＭＳキャッシュ管理ＩＤ８１１を求め、それらプールとエントリに対応するＤＢＭＳデータキャッシュ管理情報８３０のエントリを更新する。

【０１１５】実行情報ＩＤ８０１と実行情報内部通番８０６に対応するＢ－Ｔｒｅｅ索引プリフェッチ情報８２０のエントリに対しては、１セグメント分の領域を割り当て、Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０からそのエントリにおけるＤＢＭＳキャッシュ管理ＩＤ８１１を求め、プールとそのエントリに対応するＤＢＭＳデータキャッシュ管理情報８３０のエントリを更新する。このとき、すべてのプリフェッチ情報に対してプリフェッチ用のキャッシュセグメントが行き渡るようにし、プー

ル内のキャッシュセグメント量が当初に与えた値では足りなくなる場合には、シーケンシャルプリフェッチ情報８１０中の各エントリに割り当てるキャッシュ量を調節して全エントリでほぼ等しいキャッシュセグメント数が割り当てられるようにする。

【０１１６】ステップ２００９では、プリフェッチ用のプールに空きがあるか確認する。まず、実行情報ＩＤ８０１により実行情報ＩＤ管理情報８００を参照して対応するＤＢＭＳキャッシュ管理ＩＤ８１１を求めてプリフェッチ用プールの情報を取得する。そしてプリフェッチ用のプールにプリフェッチ用に未割り当てのキャッシュセグメントが存在するかどうか確認し、存在する場合には、ステップ２０１０に進み、存在しない場合にはステップ２０１４に進む。

【０１１７】ステップ２０１０では、プリフェッチ可能なデータ構造のデータを保持する物理記憶装置１８の稼働率がある閾値以上のものがあるか確認する。現在設定中の実行情報ＩＤ８０１を持つシーケンシャルプリフェッチ情報８１０とＢ－Ｔｒｅｅ索引プリフェッチ情報８２０のエントリでまだプリフェッチ用バッファが割り当てられていないものの実行情報内部通番８０６を求める。求めた実行情報ＩＤ８０１と実行情報内部通番８０６の組でＤＢＭＳ処理計画情報８０５を参照し、プリフェッチ用キャッシュがまだ割り振られていないデータ構造のデータ構造名５６１を求める。

【０１１８】現在設定中の実行計画のＤＢＭＳ名６３１と求めたデータ構造名５６１の組を用いてデータ構造物理記憶位置情報７１２を参照し、それらが記憶されているボリューム名５０１と物理記憶装置名５０２の組を用いる。求めたボリューム名５０１と物理記憶装置名５０２を用いて物理記憶装置稼動情報３２を参照し、対応するエントリ中の稼働率５９４のうち最も最近のものを求め、その値があらかじめ定められた閾値以上の場合には検索したボリューム名５０１と物理記憶装置名５０２に対応するデータ構造に対してプリフェッチを早めに実施すべきであると判断する。確認をしたすべてのデータ構造に対して、そのようなものが１つでも見つかった場合にはステップ２０１１に進み、そうでない場合にはステップ２０１２に進む。

【０１１９】ステップ２０１１では、ステップ２０１０で早めにプリフェッチを実施すべきであると判断されたデータ構造に対してプリフェッチ用のキャッシュを割り当てる。ステップ２０１０において閾値を超える稼働率を保持するボリューム名５０１と物理記憶装置名５０２の組を求めるもとになったシーケンシャプリフェッチ情報８１０またはＢ－Ｔｒｅｅ索引プリフェッチ情報８２０のエントリを求める。以下、ステップ２００８と同じ方法で対応するエントリにプリフェッチ用キャッシュを割り当て、ステップ２０１２に進む。

【０１２０】ステップ２０１２では、ステップ２００９と同じ方法でプリフェッチ用のプールに空きがあるか確認をし、空きがある場合にはステップ２０１３に進み、ない場合にはステップ２０１４に進む。

【０１２１】ステップ２０１３においては、まだプリフェッチ用キャッシュが未割り当てのプリフェッチ情報にプリフェッチ用キャッシュを割り振る処理を行う。現在設定中の実行情報ＩＤ８０１を持つシーケンシャルプリフェッチ情報８１０とＢ－Ｔｒｅｅ索引プリフェッチ情報８２０のエントリでまだプリフェッチ用バッファが割り当てられていないものの実行情報内部通番８０６を求める。その中の最も小さな実行情報内部通番８０６に対応する実行順序８０７をＤＢＭＳ処理計画情報８０５を参照して求める。

【０１２２】ＤＢＭＳ処理計画情報８０５で現在設定中の実行情報ＩＤ８０１と求めた実行順序８０７を持つエントリの実行情報内部通番８０６を求め、先に求めた現在設定中の実行情報ＩＤ８０１を持つシーケンシャルプリフェッチ情報８１０とＢ－Ｔｒｅｅ索引プリフェッチ情報８２０のエントリでまだプリフェッチ用バッファが割り当てられていないものの中でここで求めた実行情報内部通番８０６を持つエントリを選択する。それらのエントリにおいて、ステップ２００８と同じ方法で対応するエントリにプリフェッチ用キャッシュを割り当てる。

【０１２３】以下、ステップ２００９と同じ方法でプリフェッチ用のプールに空きがあるか確認をし、空きがある場合にはこの実行順序８０７をもとにしたプリフェッチ用キャッシュの割り当て処理をプリフェッチ用のプールに空きが無くなるまで繰り返し、ステップ２０１４に進む。

【０１２４】ステップ２０１４においては、これまでに割り振られたプリフェッチ用のキャッシュセグメントに対してプリフェッチアクセス先を決定する処理を行い、そのプリフェッチ先のデータがデータキャッシュ２８上に既に存在するか確認し、それに応じた処理を実施する。現在設定中の実行情報ＩＤ８０１を持つシーケンシャルプリフェッチ情報８１０のエントリでプリフェッチ用キャッシュが割り当てられたものに対して、まずデータ領域範囲８１２の領域の先頭からキャッシュの割り当て分だけプリフェッチすることとする。そして、プリフェッチポインタ８１３を割り当て分だけ進める,エントリ中の実行情報ＩＤ８０１と実行情報内部通番８０６を求め、その組を用いてＤＢＭＳ処理計画情報８０５参照して対応するデータ構造名５６１を求める。求めたデータ構造名５６１と現在設定中のＤＢＭＳ名６３１、先にデータ領域範囲８１２の先頭から割り当てたアクセス先をデータ構造ブロック通番７１６としてデータ構造物理記憶位置情報７１２を参照し対応するデータが記憶されているボリューム名５０１とボリューム論理ブロック番号５１２と物理記憶装置名５０２とその物理ブロック番号５１４を求める。

【０１２５】ボリューム名５０１とボリューム論理ブロック番号５１２を用いてキャッシュセグメント情報７２０を参照して、対応するブロックが既にデータキャッシュ２８上に存在するかどうか確認する。データが存在する場合には、プリフェッチ要求を作成せず、その分のキャッシュセグメントはプールに戻す。そのデータを保持するセグメントが再利用ＬＲＵリスト中に存在する場合には、そのキャッシュセグメントとプリフェッチ用に割り当てられたセグメントとを交換する。これらのとき、キャッシュセグメント情報７２０と、必要に応じて再利用ＬＲＵリスト情報７４３とＤＢＭＳキャッシュ管理情報８３０中の対応エントリを更新する。

【０１２６】まだデータキャッシュ２８上にデータが存在しない場合には、求めた物理記憶装置１８と物理ブロック番号５１４と読み出し先のセグメントＩＤ７２１からプリフェッチアクセス要求を作成する。このアクセス要求にはアクセス先のＤＢＭＳ名６３１とデータ構造名５６１を含めておく。

【０１２７】現在設定中の実行情報ＩＤ８０１を持つＢ－Ｔｒｅｅ索引プリフェッチ情報８２０のエントリでプリフェッチ用キャッシュが割り当てられたものに対しては、そのエントリが対応する木構造索引のルートのデータを持つ部分をプリフェッチする。エントリ中の実行情報ＩＤ８０１と実行情報内部通番８０６を求め、その組を用いてＤＢＭＳ処理計画情報８０５参照して対応するデータ構造名５６１を求め、索引を特定する。

【０１２８】求めたデータ構造名５６１と現在設定中のＤＢＭＳ名６３１からＤＢＭＳデータ記憶位置情報６２２とＤＢＭＳホスト情報６２６とＤＢＭＳホストマッピング情報６２７を参照し、木構造索引のルートデータを持つ領域のボリューム名５０１とボリューム論理ブロック番号５１２をＤＢＭＳ１１０ａ，１１０ｂに関する知識を用いて求める。

【０１２９】求めたボリューム領域の情報を用いてキャッシュセグメント情報７２０を参照して、対応するブロックが既にデータキャッシュ２８上に存在するかどうか確認する。データが存在する場合には、そのデータを解釈し、Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０のエントリ中の索引アクセス条件８２１とその他必要なマッピング情報を参照して次にアクセスするデータのボリューム名５０１とボリューム論理ブロック番号５１２を求める。

【０１３０】以下、データキャッシュ２８上に存在しないデータに当たるまでデータの解釈による次アクセス先の把握とデータキャッシュ２８上での存在確認を繰り返す。データ解釈可能範囲ですべてのデータがデータキャッシュ２８上に存在する場合にはプリフェッチに割り当てたキャッシュセグメントをプリフェッチ用のプールに戻す処理を行なう。データキャッシュ２８上に存在しないデータが見つかった場合には、そのデータのプリフェッチを考える。前述したように、Ｂ－Ｔｒｅｅ索引プリフェッチ情報８２０のエントリにプリフェッチ用に割り当てられるキャッシュセグメント数は１である。木構造索引のルートデータがデータキャッシュ２８上に存在した場合、データ解釈の結果としてアクセス先が複数の領域に分かれる可能性がある。

【０１３１】複数の領域をプリフェッチすることになった場合には、実行情報ＩＤ８０１により実行情報ＩＤ管理情報８００を参照して対応するＤＢＭＳキャッシュ管理ＩＤ８１１を求めてプリフェッチ用プールの情報を取得し、プールに要求を満たすのに十分な空きキャッシュセグメントが存在するか確認する。十分に存在する場合は必要量を、そうでない場合は利用可能分すべてをＢ－Ｔｒｅｅ索引のプリフェッチ割り当て用に回し、ＤＢＭＳデータキャッシュ管理情報８３０とキャッシュセグメント情報７２０の対応部分を更新する。すべてのプリフェッチ可能領域に対してプリフェッチ用のキャッシュを割り当てることができない場合には、データ解釈段数を距離として、ルートデータに近いものを優先してプリフェッチ用キャッシュを割り当てる。

【０１３２】プリフェッチ先が決定した後、それらのボリューム名５０１とボリューム論理ブロック番号５１２に対応する物理記憶装置名５０２とその物理ブロック番号５１４をボリューム物理記憶位置管理情報４２を参照して求め、読み出し先のセグメントＩＤ７２１を指定してプリフェッチアクセス要求を作成する。このアクセス要求にはアクセス先のＤＢＭＳ名６３１とデータ構造名５６１を含めておく。

【０１３３】このとき、シーケンシャルプリフェッチ情報８１０またはＢ－Ｔｒｅｅ索引プリフェッチ情報８２０をもとにプリフェッチ先を割り当てられたキャッシュセグメントに関して、キャッシュセグメント情報７２０中の対応するエントリのボリューム名５０１とボリューム論理ブロック番号５１２を割り当て先の領域を示すように更新し、ステータス情報７２２の価を“インバリッド”に設定する。

【０１３４】ステップ２０１５では、ステップ２０１４で作成されたプリフェッチアクセス要求をもとにプリフェッチアクセスを発行する。全てのプリフェッチアクセス要求に対して、その中のＤＢＭＳ名６３１とデータ構造名５６１からＤＢＭＳデータ構造処理優先度情報８４０を参照して処理優先度８４１を求め、この値と現在設定中の実行情報のクエリ優先度８０３をもとにあらかじめ定められたルールを用いてこのデータ構造のこの処理における処理優先度を求める。

【０１３５】求めた処理優先度を処理優先度８４１とし指示子８６７に“プリフェッチ”を指定して、プリフェッチ先の物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の読み出しキュー配列８６３中の対応するキューに追加する。キューに追加されたアクセス要求

の物理記憶装置１８へのアクセス発行制御については後述する。

【０１３６】ステップ２０１６でクエリプラン情報８８０を取得時の処理を完了する。

【０１３７】図１５にクエリプラン情報８８０に対応するクエリの完了通知を受け取った時の処理フローを示す。これは、ネットワークインターフェイス７８が受け取った場合にはネットワークインターフェイス制御部６０が、Ｉ／Ｏパスインターフェイス７０が受け取った場合にはＩ／Ｏパスインターフェイス制御部５８がキャッシュ制御部５４にクエリプラン情報８８０に対応するクエリの完了通知を受け取ったことを伝えることによりキャッシュ制御部５４が処理を開始する。ステップ２６０１で処理を開始する。このとき、完了したクエリを識別するために、ＤＢＭＳ名６３１とクエリプランＩＤ８０２が与えられる。

【０１３８】ステップ２６０２では、与えられたＤＢＭＳ名６３１とクエリプランＩＤ８０２を用いて実行情報ＩＤ管理情報８００を参照して、対応する実行情報ＩＤ８０１を求める。

【０１３９】ステップ２６０３では、求めた実行情報ＩＤ８０１を用いてシーケンシャルプリフェッチ情報８１０とＢ－Ｔｒｅｅ索引プリフェッチ情報８２０を参照し、対応するエントリを求める。対応するそれぞれのエントリ中のＤＢＭＳキャッシュ管理ＩＤ８１１を求め、それに対応するＤＢＭＳデータキャッシュ管理情報８３０中のエントリにより管理されるプリフェッチデータ管理リンクを取得し、その管理リンク中に存在するキャッシュセグメントを全て再利用ＬＲＵリストのＭＲＵ側に繋ぎ直し、それにあわせて再利用ＬＲＵリスト情報７４３とキャッシュセグメント情報７２０の対応部分を更新する。

【０１４０】そして、先に求めたＤＢＭＳキャッシュ管理ＩＤ８１１に対応するするＤＢＭＳデータキャッシュ管理情報８３０のエントリの先頭セグメントＩＤ８３１と最後尾セグメントＩＤ８３２に無効値を、領域セグメント数８３３に－１を代入してそのエントリを無効化する。その後、先に求めた実行情報ＩＤ８０１に対応するシーケンシャルプリフェッチ情報８１０とＢ－Ｔｒｅｅ索引プリフェッチ情報８２０中のエントリを全てクリアする。

【０１４１】ステップ２６０４では、先に求めた実行情報ＩＤ８０１に対応するＤＢＭＳ処理計画情報８０５中のエントリを全てクリアする。

【０１４２】ステップ２６０５では、先に求めた実行情報ＩＤ８０１を用いて実行情報ＩＤ管理情報８００中の対応するＤＢＭＳキャッシュ管理ＩＤ８１１を求め、それに対応するＤＢＭＳデータキャッシュ管理情報８３０中のエントリにより管理されるプリフェッチ用キャッシュのプールの管理リンクを取得し、その管理リンク中に存在するキャッシュセグメントを全て再利用ＬＲＵリストのＭＲＵ側に繋ぎ直し、それにあわせて再利用ＬＲＵリスト情報７４３とキャッシュセグメント情報７２０の対応部分を更新する。そして、先に求めたＤＢＭＳキャッシュ管理ＩＤ８１１に対応するするＤＢＭＳデータキャッシュ管理情報８３０のエントリをステップ２６０３と同様な方法で無効化する。その後、先に求めた実行情報ＩＤ８０１に対応する実行情報ＩＤ管理情報８００のエントリをクリアする。

【０１４３】ステップ２６０６で処理を完了する。

【０１４４】図１６に記憶装置１０がＩ／Ｏパス７１を通してＤＢホスト８０ａ，８０ｂから書き込みアクセス要求を受け取ったときの処理フローを示す。Ｉ／Ｏパス７１を通してＩ／Ｏパスインターフェイス７０に書き込み要求が到着した際、Ｉ／Ｏパスインターフェイス制御部５８がキャッシュ制御部５４に要求を伝え、キャッシュ制御部５４は処理を開始する。ステップ２１０１で処理を開始する。このとき、ＤＢホスト８０ａ，８０ｂからのアクセス先の指定は、ボリューム名５０１とボリューム論理ブロック番号５１２で示される。

【０１４５】ステップ２１０２では、データキャッシュ２８上に旧データが存在するか確認する。これは、書き込み先のボリューム名５０１とボリューム論理ブロック番号５１２を用いてキャッシュセグメント情報７２０を参照し、対応エントリを求めることにより確認する。１セグメントでも存在する場合にはステップ２１０３に進み、存在しない場合にはステップ２１０６に進む。

【０１４６】ステップ２１０３では、存在した旧データを保持するキャッシュセグメントが、プリフェッチ管理リストに存在するかどうか確認する。これは、先に求めたキャッシュセグメント情報７２０のエントリ内のステータス情報７２２とリスト情報７２３を参照し、それが属するリストの識別子と状態を調べることにより確認する。１セグメントでも存在する場合にはステップ２１０４に進み、存在しない場合にはステップ２１０５に進む。

【０１４７】ステップ２１０４では、書き込み処理に伴う、プリフェッチ用キャッシュプールの調整を行う。書き込み先として、プリフェッチ管理リストに存在するセグメント数と同数のキャッシュセグメントを再利用ＬＲＵリストから選択して確保する。そして、書き込み先の旧データを保持するセグメント毎に、そのセグメントが属するリストの識別子からそれを管理するＤＢＭＳデータキャッシュ管理情報８３０中の管理領域を示すＤＢＭＳキャッシュ管理ＩＤ８１１を求め、それによりプリフェッチ用セグメントを管理するシーケンシャルプリフェッチ情報８１０またはＴ－Ｔｒｅｅ索引プリフェッチ情報８２０の対応エントリを求め、そのエントリ中の実行情報ＩＤ８０１を用いて実行情報ＩＤ管理情報８００を参照し、プリフェッチ用プールのＤＢＭＳキャッシュ管

理ＩＤ８１１を求める。

【０１４８】求めたプールの管理情報に先に確保したキャッシュセグメントを１つ追加する。これを書き込み先の旧データを保持するセグメントすべてについて実行し、対応するキャッシュセグメント情報７２０とＤＢＭＳデータキャッシュ管理情報８３０のエントリの値を更新し、ステップ２１０５に進む。

【０１４９】ステップ２１０５では、書き込み先の旧データを保持するセグメントを現在繋がれている管理リストから外し、書き込み先キャッシュとして割り当てる。これは、ステップ２１０３で求めたセグメントが属する管理リストの識別子からどのリストに属しているかを判断し、それに応じて対応するキャッシュセグメント情報７２０とメインＬＲＵリスト情報７４１、再利用ＬＲＵリスト情報７４３、ＤＢＭＳデータキャッシュ管理情報８３０の対応エントリの値を更新し、ステップ２１０６に進む。

【０１５０】ステップ２１０６では、まだ書き込み先キャッシュセグメントが割り当てられていないものにそれを割り当てる。まだ割り当てられていないデータの保持に必要な数のキャッシュセグメントを再利用ＬＲＵリストのＬＲＵ側から確保し、再利用ＬＲＵリスト情報７４３と対応するのキャッシュセグメント情報７２０のエントリをそれにあわせて更新する。

【０１５１】ステップ２１０７では、取得したデータキャッシュにＤＢホスト８０ａ，８０ｂから送られてきたデータを転送する。

【０１５２】ステップ２１０８では、書き込み後のキャッシュ管理情報の更新処理を行う。まず、旧データを保持していたキャッシュセグメントに関して、キャッシュセグメント情報７２０の対応エントリ中のステータス情報７２２からその状態を確認し、“ダーティ”であったものの数を求める。そして、書き込み先として利用したキャッシュセグメント数から先に求めた旧データに関するダーティ数を引いた分だけダーティセグメント数カウンタ６７２の値を増加させる。

【０１５３】そして、キャッシュセグメント情報７２０中の書き込み先セグメントに対応するエントリのステータス情報７２２を“ダーティ”に設定する。その後書き込み先として利用したキャッシュセグメントをメインＬＲＵリストのＭＲＵ側に接続し、メインＬＲＵリスト情報７４１と対応するのキャッシュセグメント情報７２０のエントリをそれにあわせて更新する。

【０１５４】ステップ２１０９ではＩ／Ｏパスインターフェイス制御部５８に対してＤＢホスト８０ａ，８０ｂに書き込み処理が完了したことを報告することを依頼し、Ｉ／Ｏパスインターフェイス制御部５８はＩ／Ｏパスインターフェイス７０を利用してＩ／Ｏパス７１を介してＤＢホスト８０ａ，８０ｂに処理完了を報告する。

【０１５５】ステップ２１１０で処理を完了する。

【０１５６】図１７に記憶装置１０がＩ／Ｏパス７１を通してＤＢホスト８０ａ，８０ｂから読み出しアクセス要求を受け取ったときの処理フローを示す。Ｉ／Ｏパス７１を通してＩ／Ｏパスインターフェイス７０に読み出し要求が到着した際、Ｉ／Ｏパスインターフェイス制御部５８がキャッシュ制御部５４に要求を伝え、キャッシュ制御部５４は処理を開始する。ステップ２２０１で処理を開始する。このとき、ＤＢホスト８０ａ，８０ｂからのアクセス先は、ボリューム名５０１とボリューム論理ブロック番号５１２で示される。

【０１５７】ステップ２２０２では、データキャッシュ２８上にデータが存在するか確認する。これは、読み出し先のボリューム名５０１とボリューム論理ブロック番号５１２を用いてキャッシュセグメント情報７２０を参照し、対応エントリが存在し、かつ、ステータス情報７２２がインバリッドでないか調べることにより確認する。全要求データに関してデータキャッシュ２８上に存在する場合にはステップ２２０７に進み、そうでない場合にはステップ２２０３に進む。

【０１５８】ステップ２２０３では、アクセス先のデータを読み出すプリフェッチ処理が発行済みかどうか確認し、存在する場合にはそれをホストからの要求によるものに変更する。まず、アクセス先のボリューム名５０１とボリューム論理ブロック番号５１２を用いてボリューム物理記憶位置管理情報４２を参照し、アクセス先領域に対応する物理記憶装置名５０２と物理ブロック番号５１４を求める。

【０１５９】これをもとに、ディスクＩ／Ｏ実行管理情報８６０を参照し、その中の物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の読み出しキュー配列８６３を参照し、その中で指示子８６７が“プリフェッチ”のキューに存在するアクセス要求の中に先に求めた読み出し要求に対応するものが存在するかどうか確認する。存在した場合には、それを同じ処理優先度８３１における指示子８６７が“ホスト要求読出”のキューに繋ぎなおす。

【０１６０】ステップ２２０４では、まだ読み出し先キャッシュセグメントが割り当てられていないものにそれを割り当てる。まだ割り当てられていないデータの保持に必要な数のキャッシュセグメントを再利用ＬＲＵリスト７４３のＬＲＵ側から確保する。再利用ＬＲＵリスト情報７４３と対応するのキャッシュセグメント情報７２０のエントリをそれにあわせて更新する。キャッシュセグメント情報７２０中の対応するエントリのボリューム名５０１とボリューム論理ブロック番号５１２を読み出し先を示すように更新し、ステータス情報７２２の値を“インバリッド”に設定する。

【０１６１】ステップ２２０５では、ステップ２２０４でキャッシュセグメントが割り当てられたものに対して読み出し要求を発行する。このとき、各キャッシュセグ

メント毎に、アクセス先のボリューム名５０１とボリューム論理ブロック番号５１２を用いてボリューム物理記憶位置管理情報４２を参照し、それに対応する物理記憶装置名５０２と物理ブロック番号５１４を求める。

【０１６２】更に、アクセス先のボリューム名５０１とボリューム論理ブロック番号５１２を用いてデータ構造物理記憶位置情報７１２を参照して、その領域に対応するＤＢＭＳ１１０ａ，１１０ｂのデータ構造を示すＤＢＭＳ名６３１とデータ構造名５６１を求める。ＤＢＭＳ名６３１とデータ構造名５６１が求まった場合には、その組を用いてＤＢＭＳデータ構造処理優先情報８４０を参照し、実効処理優先度８４２を求め、このアクセスの処理優先度とする。また、求まらなかった場合には、一般データ処理優先度８４５を参照し、その値をこのアクセスの処理優先度とする。

【０１６３】求めた情報をもとに読み出しのアクセス要求を作成する。その中には、物理記憶装置名５０２と物理ブロック番号５１４とデータを読み出し先のキャッシュのセグメントＩＤ７２１を含める。データを読み出すために求めたアクセスの処理優先度を処理優先度８４１とし指示子８６７に“ホスト要求読み出し”を指定して、アクセス先の物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の読み出しキュー配列８６３中の対応するキューに追加する。キューに追加されたアクセス要求の物理記憶装置１８へのアクセス発行制御については後述する。

【０１６４】ステップ２２０６では、発行した読み出しアクセス要求が全て完了し、ＤＢホスト８０ａ，８０ｂから読み出し要求があったデータが全て揃うまで待ち、全てのデータが揃ったらステップ２２０７に進む。

【０１６５】ステップ２２０７ではＩ／Ｏパスインターフェイス制御部５８に対してＤＢホスト８０ａ，８０ｂからの読み出し要求があったデータがキャッシュ上のどこに存在するのかの情報としてセグメントＩＤ７２１の組を返す。Ｉ／Ｏパスインターフェイス制御部５８はＩ／Ｏパスインターフェイス７０を利用してＩ／Ｏパス７１を介してＤＢホスト８０ａ，８０ｂにそのデータを転送し、ＤＢホスト８０ａ，８０ｂとの間の処理を完了する。

【０１６６】ステップ２２０８では、アクセス先のデータの種類を確認する。アクセス先のボリューム名５０１とボリューム論理ブロック番号５１２を用いてデータ構造物理記憶位置情報７１２を参照して、その領域に対応するＤＢＭＳ１１０ａ，１１０ｂのデータ構造を示すＤＢＭＳ名６３１とデータ構造名５６１を求める。それが求まった場合にはＤＢデータであるとしてステップ２２０９に進み、求まらなかった場合にはステップ２２１０に進む。

【０１６７】ステップ２２０９では、ＤＢデータの読み出しアクセス後の処理を実施する。その処理の詳細は後述する。処理が完了後、ステップ２２１１に進み通してＤＢホスト８０ａ，８０ｂから読み出しアクセス要求を受け取ったときの処理を完了する。

【０１６８】ステップ２２１０では、アクセス先キャッシュセグメントをメインＬＲＵリストのＭＲＵ側に接続し、対応するキャッシュセグメント情報７２０のエントリとメインＬＲＵリスト情報７４１と必要であれば再利用ＬＲＵリスト情報７４３をそれにあわせて更新する。ステップ２２１１に進みＤＢホスト８０ａ，８０ｂから読み出しアクセス要求を受け取ったときの処理を完了する。

【０１６９】図１８と図１９はＤＢデータの読み出しアクセス後の処理の処理フローを示す。この処理を開始する際には、アクセス先のボリューム名５０１とボリューム論理ブロック番号５１２とそのデータを保持するセグメントＩＤ７２１が与えられる。ステップ２３０１で処理を開始する。

【０１７０】ステップ２３０２では、まず、アクセス先のボリューム名５０１とボリューム論理ブロック番号５１２を用いてデータ構造物理記憶位置情報７１２を参照して、その領域に対応するＤＢＭＳ１１０ａ，１１０ｂのデータ構造を示すＤＢＭＳ名６３１とデータ構造名５６１とその領域のデータ構造ブロック通番７１６を求める。求めたＤＢＭＳ名６３１とデータ構造名５６１によりデータ構造アクセス情報８２５を参照し、データアクセス情報８２６に求めたデータ構造ブロック通番７１６から識別されるアクセス先を追加する。

【０１７１】ステップ２３０３では、ＤＢＭＳ索引定義情報６２４をＤＢＭＳ名６３１とデータ構造名５６１を索引名６３５として参照し、アクセス先が木構造索引かどうか確認する。参照結果として、エントリが見つからなかった場合、あるいは索引タイプ６３６が木構造索引でない場合にはステップ２３０４に進み、索引タイプ６３６が木構造索引である場合にはステップ２３１１に進む。

【０１７２】ステップ２３０４では、データアクセス情報８２６からＤＢＭＳ名６３１とデータ構造名５６１に対応するデータをコピーし、そのデータをアクセス先のデータ構造ブロック通番７１６を用いてソートする。

【０１７３】ステップ２３０５では、アクセス先がシーケンシャルアクセスによるものかどうか確認する。その判断には、ステップ２３０４のソート結果を用いる。ソート結果を用いてアクセス先の前の領域がほぼ連続してアクセスされているかを確認し、アクセスされていたらシーケンシャルアクセスの一部とし、そうでない場合には、シーケンシャルアクセスの一部ではないとする。

【０１７４】ＤＢＭＳ１１０ａ，１１０ｂでビットマップ索引を用いた場合には、シーケンシャルアクセスに近いが、完全に連続していないアクセスが実行されることがある。そこで、シーケンシャル性の判断時には、必ず

しも完全に連続していなくともよいとする。つまり、アクセス先領域がある一定値以下の間隔で飛び飛びに存在する場合もシーケンシャルアクセスであると判断する。シーケンシャルアクセスの一部と判断される場合には、ステップ２３０６に進み、そうでない場合には、ステップ２３３９に進む。

【０１７５】ステップ２３０６では、アクセス先に対応する実行情報が存在するか確認する。先に求めたＤＢＭＳ名６３１からそれに対応する実行情報を管理する実行情報ＩＤ８０１を求め、これと先に求めたデータ構造名５６１によりＤＢＭＳ処理計画情報８０５を参照し、対応エントリを求める。そのエントリにおけるアクセスタイプ８０８を調べ、“シーケンシャル”であるものの実行情報ＩＤ８０１と実行情報内部通番８０６を求める。このとき、１つも対応するエントリが見つからない場合にはステップ２３３９に進み、そうでない場合にはステップ２３０７に進む。

【０１７６】ステップ２３０７では、シーケンシャルプリフェッチ情報８１０内に対応エントリが存在するか確認する。ステップ２３０６で求めた実行情報ＩＤ８０１と実行情報内部通番８０６の組を用いてシーケンシャルプリフェッチ情報８１０を参照し、対応するエントリの中で先に求めたアクセス先のデータ構造ブロック通番７１６がデータ領域範囲に入るものを求め、そのエントリ中のＤＢＭＳキャッシュ管理ＩＤ８１１を求め、これを用いてＤＢＭＳデータキャッシュ管理情報８３０を参照してこれにより管理されるセグメントのリストを求めて、そのリストに存在するキャッシュセグメントが保持しているデータのボリューム名５０１とブロック番号５１２を求める。

【０１７７】このプリフェッチされたデータ領域内にアクセス先のものが含まれるか確認する。このとき、アクセス先と求めたプリフェッチ済み領域が完全に一致していなくとも、その距離が一定ブロック数以下であれば含まれると判断する。この距離は、データ構造ブロック通番７１６の差で判断し、プリフェッチ済み領域の対応値はボリューム名５０１とブロック番号５１２を用いてデータ構造物理記憶位置情報７１２を参照することにより求めることができる。この判断により、含まれると判断された場合には、そのシーケンシャルプリフェッチ情報８１０のエントリがアクセス先に対応するエントリと判断してステップ２３３０に進み、そのようなものが見つからなかった場合には、ステップ２３０８に進む。

【０１７８】ステップ２３０８では、アクセス先データを保持するキャッシュセグメントを再利用ＬＲＵリストのＭＲＵ側に接続し、再利用ＬＲＵリスト情報７４３と対応するのキャッシュセグメント情報７２０のエントリそして必要に応じてメインＬＲＵ情報７４１をそれにあわせて更新する。

【０１７９】ステップ２３０９では、シーケンシャルプリフェッチ情報にアクセス先におけるシーケンシャルアクセスの情報を設定する。先に求めた、アクセス先に対応する実行情報ＩＤ８０１と実行情報内部通番８０６の組をもとに新規にシーケンシャルプリフェッチ情報８１０を設定する。なお、アクセス先に対応する実行情報ＩＤ８０１と実行情報内部通番８０６の組が複数存在する場合には、どれか１つを適当に選択する。

【０１８０】アクセス先が他のシーケンシャルプリフェッチ情報８１０内の既に設定されている同じ実行情報ＩＤ８０１と実行情報内部通番８０６を持つエントリのデータ領域範囲８１２に含まれる場合には、既存のエントリのデータ領域範囲８１２を現在確認中のアクセス先より前の部分のみとし、後半部分を今回新たに設定するシーケンシャルプリフェッチ情報８１０のエントリのデータ領域範囲８１２に割り当てる。そうでない場合には、データ領域範囲８１２は現在のアクセス先にあらかじめ定められた動的プリフェッチ領域拡張用のブロック数を足したものに設定する。このとき、他の同じ実行情報ＩＤ８０１と実行情報内部通番８０６を持つ既存エントリと重なる部分がある場合は現在設定中のデータ領域範囲８１２において重なり部分を切り捨てる。

【０１８１】プリフェッチポインタ８１３を現在のアクセス先の次のブロック番号に設定し、アクセス状況８１４には“実行”を入れる。更に、ＤＢＭＳデータキャッシュ管理情報８３０中の空きエントリを捜し、そのＤＢＭＳキャッシュ管理ＩＤ８１１を設定する。そして、ＤＢＭＳデータキャッシュ管理情報８３０中の選択したＤＢＭＳキャッシュ管理ＩＤを持つエントリの領域セグメント数８３３を０に設定し、ステップ２３３２に進む。

【０１８２】ステップ２３３０では、シーケンシャルプリフェッチ情報８１０のアクセス先に対応するエントリにおいて、シーケンシャルアクセス時に現在確認中のアクセス先より前にアクセスされるべき領域のデータをプリフェッチしていたものと確認中のアクセス先のものを実行情報のプリフェッチ用キャッシュのプールに戻す処理を行う。ステップ２３０７で既にそのエントリに対応するＤＢＭＳキャッシュ管理ＩＤ８１１とそれにより管理されるセグメントのリストは求まっているので、それからプールに戻すキャッシュセグメントを求める。

【０１８３】また、シーケンシャルプリフェッチ情報８１０の対応するエントリ中の実行情報ＩＤ８０１を用いて実行情報ＩＤ管理情報８００を参照して対応する実行情報のプリフェッチ用キャッシュのプールを管理するＤＢＭＳキャッシュ管理ＩＤ８１１を求め、先に求めたプールに戻すキャッシュセグメントをプールに追加する。そして、それにあわせて対応するＤＢＭＳデータキャッシュ管理情報８３０とキャッシュセグメント情報７２０のエントリを更新する。

【０１８４】ステップ２３３１では、対応するシーケンシャルプリフェッチ情報８１０のエントリを更新する。

アクセス先がプリフェッチプリフェッチポインタ８１３を追い越したらその値をアクセス先に１を加算したものに更新する。また、データ領域範囲８１２の拡張を、アクセス先がシーケンシャルプリフェッチ情報８１０のエントリ内のデータ領域範囲８１２の終端から一定の距離の場所まで進み、かつ、それに連続したデータ領域範囲８１２と同じ実行情報ＩＤ８０１と実行情報内部通番８０６を持つエントリがシーケンシャルプリフェッチ情報８１０に存在しない場合に実行する。この場合、あらかじめ定められた動的プリフェッチ領域拡張用のブロック数を最大値として他のエントリと重ならない分量をデータ領域範囲８１２の終端に足し、ステップ２３３２に進む。

【０１８５】ステップ２３３２においては、現在確認中の実行情報のプリフェッチ用キャッシュのプールの残量を確認する。シーケンシャルプリフェッチ情報８１０の対応するエントリ中の実行情報ＩＤ８０１を用いて実行情報ＩＤ管理情報８００を参照して対応する実行情報のプリフェッチ用キャッシュのプールを管理するＤＢＭＳキャッシュ管理ＩＤ８１１を求め、ＤＢＭＳデータキャッシュ管理情報８３０を参照してそれに対応する領域セグメント数８３３を求める。その値があらかじめ定められた閾値以上の場合にはステップ２３３３に進み、そうでない場合にはステップ２３４２に進み処理を完了する。

【０１８６】ステップ２３３３では、現在のアクセス先に対応するシーケンシャルプリフェッチ情報８１０により管理される領域にプリフェッチすべきデータが存在するかどうか確認する。これは、エントリのプリフェッチポインタ８１３がデータ領域範囲８１２内に存在するかどうかを確認することにより判断でき、存在する場合には、プリフェッチすべきデータがあるとしてステップ２３３７に進み、存在しない場合にはプリフェッチすべきデータが存在しないとしてステップ２３３４に進む。

【０１８７】ステップ２３３４では、次にプリフェッチを実施すべきシーケンシャルプリフェッチ情報８１０内のエントリの検索と決定を行う。現在確認中のエントリに対応するシーケンシャルプリフェッチ情報８１０のエントリが保持する実行情報ＩＤ８０１によりシーケンシャルプリフェッチ情報８１０のエントリを検索し、そのＤＢＭＳキャッシュ管理情報ＩＤ８１１を求め、それを用いてＤＢＭＳデータキャッシュ管理情報８３０を参照してプリフェッチされたセグメント数が０、かつ、アクセス状況８１４が未実行なものを選択する。

【０１８８】選択されたシーケンシャルプリフェッチ情報８１０のエントリ中で実行情報内部通番８０６が最も小さいものを次にプリフェッチを行うエントリとして選択する。このとき、複数エントリが同じ実行情報内部通番８０６を持つ可能性があるが、その場合には、対象の中でデータ領域範囲８１２が示す値が最も小さなものを選択する。

【０１８９】ステップ２３３５では、ステップ２３３４での次のプリフェッチ対象となるシーケンシャルプリフェッチ情報８１０のエントリの選択が成功したかを判断する。選択に成功した場合にはステップ２３３６に進み、既にプリフェッチすべきものが存在せず選択に失敗した場合にはステップ２３４２に進み処理を完了する。

【０１９０】ステップ２３３６では、プリフェッチ用の領域を割り当て、プリフェッチコマンドを発行する。まず、シーケンシャルプリフェッチ情報８１０の対応するエントリ中の実行情報ＩＤ８０１を用いて実行情報ＩＤ管理情報８００を参照し、対応する実行情報のＤＢＭＳキャッシュ管理ＩＤ８１１を求め、ＤＢＭＳデータキャッシュ管理情報８３０を参照して、プリフェッチ用キャッシュのプールを管理するリストを取得し、そこからプリフェッチ用に割り当てるキャッシュセグメントを確保する。プリフェッチ実行量はあらかじめ定められた量であるとする。

【０１９１】この確保したキャッシュセグメントをシーケンシャルプリフェッチ情報８１０の対応するＤＢＭＳデータキャッシュ管理ＩＤ８１１で示されるＤＢＭＳデータキャッシュ管理情報８３０で管理されるプリフェッチデータ管理リストに繋ぎ、ＤＢＭＳデータキャッシュ管理情報８３０のプリフェッチ用プールとプリフェッチデータ管理リストに対応する部分とキャッシュセグメント情報７２０中の対応する部分を更新する。

【０１９２】確保したキャッシュセグメントにプリフェッチ先を割り当て、キャッシュセグメント情報７２０中の対応するエントリのボリューム名５０１とボリューム論理ブロック番号５１２を割り当て先の領域を示すように更新し、ステータス情報７２２の価を“インバリッド”に設定する。アクセス先は、プリフェッチ対象となったシーケンシャルプリフェッチ情報８１０の対応するエントリのプリフェッチポインタ８１３から連続して割り当てられたキャッシュ量分の領域である。プリフェッチポインタ８１３はデータ構造ブロック通番７１６に対応している。そして、プリフェッチポインタ８１３を割り当て分だけ進める。シーケンシャルプリフェッチ情報８１０の対応するエントリの実行情報ＩＤ８０１から実行情報ＩＤ管理情報８００を参照して対応するＤＢＭＳ名６３１を求め、エントリの実行情報ＩＤ８０１と実行情報内部通番８０６を用いてＤＢＭＳ処理計画情報８０５を参照してデータ構造名５６１を求め、これらの値を用いてデータ構造物理記憶位置情報７１２を参照してアクセス先のアクセスする物理記憶装置１８上の領域である物理記憶装置名５０２と物理ブロック番号５１４を求める。また、ＤＢＭＳ名６３１とデータ構造名５６１を用いてＤＢＭＳデータ構造処理優先情報８４０を参照し、実効処理優先度８４２を求め、このアクセスの処理優先度とする。

【０１９３】割り当てたキャッシュセグメントを用いてプリフェッチアクセス要求を作成する。その中には、物理記憶装置名５０２と物理ブロック番号５１４とデータを読み出し先のキャッシュのセグメントＩＤ７２１を含める。求めたアクセスの処理優先度を処理優先度８４１とし指示子８６７に“プリフェッチ”を指定して、アクセス先の物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の読み出しキュー配列８６３中の対応するキューに追加する。キューに追加されたアクセス要求の物理記憶装置１８へのアクセス発行制御については後述する。その後、ステップ２３４２に進み処理を完了する。

【０１９４】ステップ２３３７では、アクセス先に対応するシーケンシャルプリフェッチ情報８１０のエントリにおける既プリフェッチデータ量を確認する。この値は、シーケンシャルプリフェッチ情報８１０の対応するＤＢＭＳデータキャッシュ管理ＩＤ８１１によりＤＢＭＳデータキャッシュ管理情報８３０を参照し、対応する領域セグメント数８３３の値を求める。この値が既プリフェッチデータ量に対応し、この値がある閾値未満の場合にはステップ２３３８に進み、閾値以上である場合には、ステップ２３４２に進み処理を完了する。

【０１９５】ステップ２３３８では、アクセス先に対応するシーケンシャルプリフェッチ情報８１０のエントリで管理される領域をプリフェッチの実行対象に設定し、ステップ２３３６に進む。

【０１９６】ステップ２３１１では、現アクセス先データの中身を解釈することによりプリフェッチが可能かどうか確認する。先に求めたＤＢＭＳ名６３１からそれに対応する実行情報を管理する実行情報ＩＤ８０１を求め、これと先に求めたデータ構造５６１によりＤＢＭＳ処理計画情報８０５を参照し、対応エントリを求める。ここで対応エントリが求まらない場合には、プリフェッチ不可能としてステップ２３３９に進む。求まった場合、そのエントリの実行情報ＩＤ８０１と実行情報内部通番８０６を用いてＢ－Ｔｒｅｅ索引プリフェッチ情報８２０を参照し、対応エントリを求める。ここで対応エントリが求まらない場合には、プリフェッチ不可能としてステップ２３３９に進む。求まった場合には、ステップ２３１２に進む。なお、複数のエントリが求まる可能性がある。

【０１９７】ステップ２３１２では、読み出しデータからプリフェッチ先を求める処理を行う。ステップ２３１１において求めたＢ－Ｔｒｅｅ索引プリフェッチ情報８２０の全てのエントリにおける索引アクセス条件８２１を求め、アクセス先のデータの解釈をして索引アクセス条件８２１によりアクセスされるデータのボリューム名５０１とボリューム論理ブロック番号５１２を求める。このアクセス先は複数存在する可能性がある。

【０１９８】次アクセス先のボリューム名５０１とボリューム論理ブロック番号５１２からキャッシュセグメント情報７２０を参照し、次データがデータキャッシュ２８上に存在するかどうか確認する。次データが存在する場合には、再帰的に解釈処理を続ける。アクセス先が複数存在する場合には、それぞれで次アクセス先データのデータキャッシュ２８上の存在確認とデータ解釈処理を実施する。データ解釈可能範囲ですべてのデータがデータキャッシュ２８上に存在する場合にはプリフェッチ先なしとしてステップ２３３９に進み、そうでない場合には２３１３に進む。

【０１９９】ステップ２３１３では、現在確認中の実行情報のプリフェッチ用キャッシュのプールの残量を確認する。ステップ２３１１で求めた実行情報ＩＤ８０１を用いて実行情報ＩＤ管理情報８００を参照して対応する実行情報のプリフェッチ用キャッシュのプールを管理するＤＢＭＳキャッシュ管理ＩＤ８１１を求め、ＤＢＭＳデータキャッシュ管理情報８３０を参照して、プリフェッチ用キャッシュのプールを管理するリストを取得し、その中の管理情報である領域セグメント数８３３を求める。その値が０でなかったら２３１５に進み、その値が０であったらステップ２３３９に進む。

【０２００】ステップ２３１４では、ステップ２３１２で求めたプリフェッチ先に対してプリフェッチ用の領域を割り当て、プリフェッチコマンドを発行する。まず、ステップ２３１３で求めたプリフェッチ用キャッシュのプールを管理するリストからプリフェッチ用に割り当てるキャッシュセグメントを確保する。確保する量はステップ２３１２により求まったアクセス先に対応する量である。領域が不足する場合には可能な量全てを確保する。

【０２０１】この確保したキャッシュセグメントをＢ－Ｔｒｅｅ索引プリフェッチ情報８２０の対応するＤＢＭＳデータキャッシュ管理ＩＤ８１１で示されるＤＢＭＳデータキャッシュ管理情報８３０のエントリで管理されるプリフェッチデータ管理リストに繋ぎ、ＤＢＭＳデータキャッシュ管理情報８３０のプリフェッチ用プールとプリフェッチデータ管理リストに対応する部分とキャッシュセグメント情報７２０中の対応する部分を更新する。確保したキャッシュセグメントにプリフェッチ先を割り当て、キャッシュセグメント情報７２０中の対応するエントリのボリューム名５０１とボリューム論理ブロック番号５１２を割り当て先の領域を示すように更新し、ステータス情報７２２の値を“インバリッド”に設定する。

【０２０２】このとき、キャッシュセグメントが要求量全てを確保できなかったときには、データ解釈段数を距離として、データ解釈を開始したデータに近いものに優先してプリフェッチ用キャッシュを割り当てる。アクセス先に関しては、ボリューム名５０１とボリューム論理ブロック番号５１２が求まっているため、データ構造物

理記憶位置情報７１２を参照して、アクセス先の物理記憶装置１８上の領域である物理記憶装置名５０２と物理ブロック番号５１４とこのデータが属するＤＢＭＳ名６３１とデータ構造名５６１を求める。

【０２０３】求めたＤＢＭＳ名６３１とデータ構造名５６１を用いてＤＢＭＳデータ構造処理優先情報８４０を参照し、実効処理優先度８４２を求め、このアクセスの処理優先度とする。割り当てたキャッシュセグメントを用いてプリフェッチアクセス要求を作成する。その中には、物理記憶装置名５０２と物理ブロック番号５１４とデータを読み出し先のキャッシュのセグメントＩＤ７２１を含める。

【０２０４】求めたアクセスの処理優先度を処理優先度８４１とし指示子８６７に“プリフェッチ”を指定して、アクセス先の物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の読み出しキュー配列８６３中の対応するキューに追加する。キューに追加されたアクセス要求の物理記憶装置１８へのアクセス発行制御については後述する。その後、ステップ２３３９に進む。

【０２０５】ステップ２３３９においては、現在確認中のアクセス先データを保持するキャッシュセグメントがプリフェッチデータ管理リストに存在するか確認する。データを保持するキャッシュのセグメントＩＤ７２１を用いてキャッシュセグメント情報７２０を参照し、対応するリスト情報７２３を参照してどの管理リストにより管理されているかを判断する。プリフェッチデータの管理リストに存在する場合にはステップ２３４０に進み、そうでない場合にはステップ２３４１に進む。

【０２０６】ステップ２３４０においては、現在確認中のアクセス先データを保持するキャッシュセグメントを、そのキャッシュセグメントを管理するプリフェッチデータ管理リストを含む実行情報におけるプリフェッチ用キャッシュのプールに戻す処理を行う。ステップ２３３９で対応するプリフェッチデータ管理リストが求まっているため、それに対応するＤＢＭＳキャッシュ管理ＩＤ８１１を求める。この価をもとに、プリフェッチデータ管理リストを保持しているシーケンシャルプリフェッチ情報８１０またはＢ－Ｔｒｅｅプリフェッチ情報８２０のエントリを求める。どちらにも属していない場合には、既にプリフェッチ用キャッシュのプールに存在しているので以下のリストの変更処理は行わない。

【０２０７】求まったシーケンシャルプリフェッチ情報８１０またはＢ－Ｔｒｅｅプリフェッチ情報８２０のエントリにおける実行情報ＩＤ８０１を用いて実行情報ＩＤ管理情報８００を参照して、キャッシュセグメントを戻すプリフェッチ用キャッシュのプールのリストを管理するＤＢＭＳキャッシュ管理ＩＤ８１１を求める。現在確認中のアクセス先データを保持するキャッシュセグメントを現在管理するプリフェッチデータ管理リストから外し、それをプリフェッチ用キャッシュのプールのリストに繋ぎ、それに対応してＤＢＭＳデータキャッシュ管理情報８３０のプリフェッチ用プールとプリフェッチデータ管理リストに対応する部分とキャッシュセグメント情報７２０中の対応する部分を更新する。これらの処理が完了後、ステップ２３４２に進み処理を完了する。

【０２０８】ステップ２３４１においては、現在確認中のアクセス先データを保持するキャッシュセグメント現在の管理リストから一旦外してそれをメインＬＲＵリストのＭＲＵ側に接続し、対応するのキャッシュセグメント情報７２０のエントリとメインＬＲＵリスト情報７４１と必要であれば再利用ＬＲＵリスト情報７４３をそれにあわせて更新する。この処理が完了後、ステップ２３４２に進み処理を完了する。

【０２０９】図２０にディスクＩ／Ｏ実行管理情報８６０を利用した物理記憶装置１８へのアクセス処理を行うバックグラウンド処理の処理フローを示す。この処理は、物理記憶装置１８毎に実行され、記憶装置１０が動作を開始すると同時にディスクコントローラ制御部５２で開始され、その処理は無限ループとなっている。ステップ２４０１で処理を開始する。このとき、対象なる物理記憶装置１８の識別子である物理記憶装置名５０２が指定される。

【０２１０】ステップ２４０２では、ディスクＩ／Ｏ実行管理情報８６０の対象とする物理記憶装置名５０２に対するＩ／Ｏ実行管理情報８６１中の処理優先度別処理残数カウンタ８６６の中の処理残りカウント値８６８を初期化する。ここでは、処理優先度情報８５０を参照し、処理優先度８４１毎の割り当てＩ／Ｏ比率８５１を求め、それにあらかじめ与えられた比率を乗じたものを処理優先度別処理残数カウンタ８６６の中の処理残りカウント値８６８の対応する部分に代入する。初期化後、ステップ２４０３に進む。

【０２１１】ステップ２４０３では、キャッシュセグメント利用管理情報７４０中のダーティセグメント数カウンタ７４６の値を確認する。その価があらかじめ定められたある閾値以上である場合には、ステップ２４０４に進み、そうでない場合には、ステップ２４０７に進む。

【０２１２】ステップ２４０４では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の書き込みキュー８６４に書き込み要求が存在するか確認する。存在する場合には、ステップ２４０５に進み、存在しない場合にはステップ２４０７に進む。

【０２１３】ステップ２４０５では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の書き込みキュー８６４の先頭に存在する書き込み要求を１つ取り出し、対象とする物理記憶装置名５０２を持つ物理記憶装置１８への書き込み処理を実行する。その書き込み処理の完了を待ち、処理完了後、キャッシュセグメント情報７２０の更新を行う。対象とするセグメントＩＤ７２１はアクセス要求中に含まれているので、キ

ャッシュセグメント情報７２０中のそれに対応するエントリのステータス情報７２２を“ノーマル”に設定する。その後、ステップ２４０３に進む。なお、ここで、複数の書き込み要求を同時に実行してもよい。

【０２１４】ステップ２４０７では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中のホスト要求読出アクセスキュー中にアクセス要求が存在するか確認する。このとき、処理優先度別処理残数カウンタ８６６を考慮する。まず処理優先度別処理残数カウンタ８６６中の処理残りカウント値が０でない最も高い処理優先度８４１を求める。求めた処理優先度８４１と指示子８６７“ホスト要求読出”を用いて読み出しキュー配列８６３を参照し、対応するアクセスキューにアクセス要求が存在したらその処理優先度８４１を記憶してステップ２４１１に進む。存在しない場合、処理優先度別処理残数カウンタ８６６中の処理残りカウント値が０でない２番目に高い処理優先度８４１を求め、同じようにアクセス要求を確認し、存在したらその処理優先度を記憶してステップ２４１１に進む。

【０２１５】以下同様に処理優先度別処理残数カウンタ８６６中の処理残りカウント値を考慮しながら処理優先度８４１とその価におけるアクセス要求が存在するか確認する処理を繰り返し、存在する場合にはそのときの処理優先度を記憶してステップ２４１１に進み、存在しない場合にはステップ２４０８に進む。

【０２１６】ステップ２４０８では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中のプリフェッチアクセスキュー中にアクセス要求が存在するか確認する。このとき、処理優先度別処理残数カウンタ８６６を考慮する。まず処理優先度別処理残数カウンタ８６６中の処理残りカウント値が０でない最も高い処理優先度８４１を求める。

【０２１７】求めた処理優先度８４１と指示子８６７“プリフェッチ”を用いて読み出しキュー配列８６３を参照し、対応するアクセスキューにアクセス要求が存在したらその処理優先度８４１を記憶してステップ２４１２に進む。存在しない場合、以下同様に処理優先度別処理残数カウンタ８６６中の処理残りカウント値を考慮しながら処理優先度８４１とその値におけるアクセス要求が存在するか確認する処理を繰り返し、存在する場合にはそのときの処理優先度を記憶してステップ２４１２に進み、存在しない場合にはステップ２４０９に進む。

【０２１８】ステップ２４０９では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の書き込みキュー８６４に書き込み要求が存在するか確認する。存在する場合には、ステップ２４０５に進み、存在しない場合にはステップ２４１０に進む。

【０２１９】ステップ２４１０に到着するときは、アクセス要求が全く存在しないときか、あるいは、処理優先度別処理残数カウンタ８６６を考慮するために存在するアクセス要求を実行できなかったときである。ここでは、読み出しキュー配列８６３とアクセス要求が書き込みキュー８６４にアクセス要求が全く存在しないか確認する。１つでも存在する場合には、即座にステップ２４０２に進む。全く存在しない場合には、新たにアクセス要求が発行されるまでこのステップで待ち、アクセス要求が発行された時点で、ステップ２４０２に進む。

【０２２０】ステップ２４１１では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の読み出しキュー配列８６３をステップ２４０７で記憶した処理優先度８４１と指示子８６７として“ホスト要求読出”を参照し、対応するアクセスキューに存在するアクセス要求を１つ取り出し、対象とする物理記憶装置名５０２を持つ物理記憶装置１８からの読み出し処理を実行する。

【０２２１】その読み出し処理の完了を待ち、処理完了後、キャッシュセグメント情報７２０の更新を行う。対象とするセグメントＩＤ７２１はアクセス要求中に含まれているので、キャッシュセグメント情報７２０中のそれに対応するエントリのステータス情報７２２を“ノーマル”に設定する。その後、ステップ２４１３に進む。

【０２２２】なお、ここで、複数の読み出し要求を同時に実行してもよい。ただし、その上限値は処理優先度別処理残数カウンタ８６６のステップ２４０７で記憶された処理優先度８４１に対応する処理残りカウント値８６８の値である。

【０２２３】ステップ２４１２では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の読み出しキュー配列８６３をステップ２４０８で記憶された処理優先度８４１と指示子８６７として“プリフェッチ”を参照し、対応するアクセスキューに存在するアクセス要求を１つ取り出し、対象とする物理記憶装置名５０２を持つ物理記憶装置１８からの読み出し処理を実行する。

【０２２４】その読み出し処理の完了を待ち、処理完了後、キャッシュセグメント情報７２０の更新を行う。対象とするセグメントＩＤ７２１はアクセス要求中に含まれているので、キャッシュセグメント情報７２０中のそれに対応するエントリのステータス情報７２２を“ノーマル”に設定する。

【０２２５】その後、ステップ２４１３に進む。なお、ここで、複数の読み出し要求を同時に実行してもよい。ただし、その上限値は処理優先度別処理残数カウンタ８６６のステップ２４０８で記憶された処理優先度８４１に対応する処理残りカウント値８６８の値である。

【０２２６】ステップ２４１３では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の処理優先度別処理残数カウンタ８６６の処理残りカウント値８６８の更新を行う。ステップ２４１１もしくは２４１２で実行した読み出しアクセス数を、その読み出

し要求が存在していた処理優先度８４１に対応する処理残りカウント値８６８から減する。

【０２２７】ステップ２４１４では、対象とする物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の処理優先度別処理残数カウンタ８６６の処理残りカウント値８６８の確認を行う。対象である処理優先度別処理残数カウンタ８６６中の全ての処理残りカウント値８６８の値が０以下の場合には、ステップ２４０２に進み、そうでない場合にはステップ２４０３に進む。

【０２２８】図２１にキャッシュ制御部５４がバックグラウンドで実行する周期処理の処理フローを示す。この処理は、ダーティセグメント数を一定値以下に、再利用ＬＲＵリストに存在するセグメント数を一定値以上に保つための処理で、記憶装置１０が動作を開始すると同時に例えば１秒周期等の一定周期で処理が開始される。ステップ２５０１で処理が開始される。

【０２２９】ステップ２５０２では、キャッシュセグメント利用管理情報７４０中のダーティセグメント数カウンタ７４６の値を確認する。その値があらかじめ定められた閾値以上の場合にはステップ２５０３に進み、閾値未満の場合にはステップ２５０４に進む。

【０２３０】ステップ２５０３では、ダーティセグメント数カウンタ７４６の価が前記閾値未満になるようにキャッシュセグメント情報７２０中のステータス情報７２２が“ダーティ”状態のキャッシュセグメントの一部のデータを物理記憶装置１８に対して書き込むアクセス要求を発行する処理を行う。ここで、書き込むセグメント数は、ダーティセグメント数カウンタ７４６から前記閾値を引いた値にあらかじめ定めてある一定値を加えたものとする。

【０２３１】メインＬＲＵリストのＬＲＵ側からキャッシュセグメント情報７２０中のステータス情報７２２が“ダーティ”であるキャッシュセグメントを先に求めた書き込みセグメント数だけ選択し、そのステータス情報７２２を“ライト”に設定する。そして、求めたキャッシュセグメントのセグメントＩＤ７２１とそのセグメントに記憶されているデータのボリューム名５０１とボリューム論理ブロック番号５１２を求める。

【０２３２】求めたボリューム名５０１とボリューム論理ブロック番号５１２から対応する物理記憶装置名５０２と物理論理番号５１４を求め、更にそのデータを保持するキャッシュのセグメントＩＤ７２１を指定して書き込みアクセス要求を作成し、ディスクＩ／Ｏ実行管理情報８６０中のアクセス先の物理記憶装置名５０２に対応するＩ／Ｏ実行管理情報８６１中の書き込みキュー８６４中に追加する。その後、作成した書き込みアクセス要求の数だけダーティセグメント数カウンタ７４６の値を減じ、ステップ２５０４に進む。

【０２３３】ステップ２５０４では、キャッシュセグメント利用管理情報７４０中の再利用ＬＲＵリスト情報７４３中に記憶されているそのリストに存在するセグメント数を確認する。その価があらかじめ定められた閾値未満の場合にはステップ２５０５に進み、閾値以上の場合にはステップ２５０６に進み、処理を完了する。

【０２３４】ステップ２５０５では、メインＬＲＵリストに存在するセグメントの一部を再利用ＬＲＵリストに移動し、再利用ＬＲＵリストに存在するセグメント数を増加させる処理を行う。ここで、移動させるセグメント数は、再利用ＬＲＵリスト情報７４３中のリストに存在するセグメント数から前記閾値を引いた値にあらかじめ定めてある一定値を加えたものとする。

【０２３５】メインＬＲＵリストのＬＲＵ側からキャッシュセグメント情報７２０中のステータス情報７２２が“ノーマル”であるキャッシュセグメントを先に求めた移動させるセグメント数だけ選択してメインＬＲＵリストから外し、それらを再利用ＬＲＵリストのＭＲＵ側に繋ぎ直す。そして、対応するキャッシュセグメント情報７２０のエントリとメインＬＲＵリスト情報７４１と再利用ＬＲＵリスト情報７４３をそれにあわせて更新する。その後、ステップ２５０６に進み、処理を完了する。＜第二の実施の形態＞本実施の形態では、ＤＢＭＳが実行される計算機と、ファイルを管理単位とし、データキャッシュを保持する記憶装置が接続された計算機システムにおいて、記憶装置がＤＢＭＳに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報、ＤＢＭＳで実行されるクエリの実行プラン情報、ＤＢの処理優先順位情報を取得し、それらを用いて記憶装置がより好ましいアクセス性能を提供する。記憶装置は、ＤＢＭＳに関する情報、記憶装置外におけるデータの記憶位置のマッピングに関する情報、ＤＢＭＳで実行されるクエリの実行プランを利用することにより、ＤＢＭＳがこれからどのデータをどの順序でどのようにアクセスするかを把握することができる。

【０２３６】そこで、この把握したアクセス方法に関する情報を利用して、あらかじめ利用される可能性が高いデータを記憶装置上のデータキャッシュ上に用意しておくことにより、ＤＢＭＳに対してより高いアクセス性能を提供する。また、ＤＢの処理優先順位情報を利用し、処理優先度が高いＤＢのデータに対して、記憶装置が保持する物理記憶装置へアクセスを優先的に実施したり、また、データキャッシュの利用量をより多く割り当てたりすることにより、処理優先度が高いＤＢのデータに対するアクセス性能を向上させる。

【０２３７】図２２は、本発明の第二の実施の形態における計算機システムの構成図である。図示されたように、本発明の第二の実施の形態は本発明の第一の実施の形態と以下の点が異なる。

【０２３８】本実施の形態においてはＩ／Ｏパスインターフェイス７０、Ｉ／Ｏパス７１、Ｉ／Ｏパススイッチ７２が存在せず、記憶制御装置１０ｂとＤＢホスト８０

ｃ，８０ｄはネットワーク７９を介してのみ接続される。記憶装置１０はファイルを単位としたデータ記憶管理を行う記憶装置１０ｂに変更される。そのため、物理記憶装置稼動情報３２、データキャッシュ管理情報３４、処理優先度付ディスクＩ／Ｏ管理情報３６、ＤＢＭＳ実行情報３８、ＤＢＭＳデータ情報４０、ボリューム物理記憶位置管理情報４２がそれぞれ物理記憶装置稼動情報３２ｂ、データキャッシュ管理情報３４ｂ、処理優先度付ディスクＩ／Ｏ管理情報３６ｂ、ＤＢＭＳ実行情報３８ｂ、ＤＢＭＳデータ情報４０ｂ、ファイル記憶管理情報４２ｂに変更される。

【０２３９】ＤＢホスト８０ｃ，８０ｄで実行されるＯＳ１００ではボリュームマネージャ１０２、ファイルシステム１０４が削除されその代わりに記憶装置１０ｂが提供するファイルをアクセスするための機能を有するネットワークファイルシステム１０４ｂが追加され、ＯＳ１００が保持するマッピング情報１０６がネットワークマウント情報１０６ｂへ変更される。

【０２４０】記憶装置１０はファイルを管理単位とする記憶装置１０ｂに変更される。ＤＢホスト８０ｃ，８０ｄからのアクセスもＮＦＳ等のファイルをベースとしたプロトコルで実施される。記憶装置１０におけるボリュームの役割は、記憶装置１０ｂにおいてはファイルもしくはファイルを管理するファイルシステムとなり、そのファイルの記憶位置管理情報がファイル記憶管理情報４２ｂである。１つの記憶装置１０ｂの中に複数のファイルシステムが存在しても構わない。物理記憶装置１８の稼動情報はボリュームを単位とした取得からファイルシステムまたはファイルを単位とした取得に変更する。記憶装置１０ｂ内にファイルシステムが存在する場合でもデータの移動機能を実現可能である。

【０２４１】図２３はＤＢホスト８０ｃ，８０ｄのＯＳ１００内に記憶されているネットワークマウント情報１０６ｂを示す。ネットワークマウント情報１０６ｂは、記憶装置１０ｂから提供され、ＤＢホスト８０ｃ，８０ｄにおいてマウントされているファイルシステムの情報で、ファイルシステムの提供元記憶装置とそのファイルシステムの識別子である記憶装置名５８３とファイルシステム名１００１、そして、そのファイルシステムのマウントポイントの情報であるマウントポイント１０３１の組を保持する。

【０２４２】図２４は記憶装置１０ｂ内に保持されるファイル記憶管理情報４２ｂを示す。図４のボリューム物理記憶位置管理情報４２からの変更点は、ボリュームの識別子であるボリューム名５０１がファイルの識別子となるファイルシステム名１００１とファイルパス名１００２に、ボリューム内のデータ領域を示すボリューム論理ブロック番号５１２がファイルブロック番号１００３変更されるものである。

【０２４３】図２５に記憶装置１０ｂ内に保持される物理記憶装置稼動情報３２ｂを示す。図５の物理記憶装置稼動情報３２からの変更点は、稼動情報取得単位がボリュームからファイルシステムに変更されたため、ボリューム名５０１の部分がファイルシステム名１００１に変更されたことである。また、稼動情報取得単位をファイルとしてもよく、このときはボリューム名５０１の部分がファイルシステム名１００１とファイルパス名１００２に変更される。図２６に記憶装置１０ｂ内に保持されているＤＢＭＳデータ情報４０ｂを示す。図６のＤＢＭＳデータ情報４０からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたためデータ構造物理記憶位置情報７１２に修正が加えられ、データ構造物理記憶位置情報７１２ｂに変更されたことと、ＤＢＭＳスキーマ情報７１１中のＤＢＭＳホストマッピング情報６２７中のマッピング情報６４８が保持するデータがＤＢホストにおけるマッピング情報１０６からネットワークマウント情報１０６６に変更されたことである。

【０２４４】図２７にＤＢＭＳデータ情報４０ｂ中に含まれるデータ構造物理記憶位置情報７１２ｂを示す。図８のデータ構造物理記憶位置情報７１２からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、ボリューム名５０１とボリューム論理ブロック番号５１２の部分がファイルシステム名１００１とファイルパス名１００２とファイルブロック番号１００３に変更されたことである。この情報は、ＤＢＭＳスキーマ情報７１１内のＤＢＭＳデータ記憶位置情報６２２とＤＢＭＳホストマッピング情報６２７とファイル記憶管理情報４２ｂを参照して、対応する部分を組み合わせることにより作成する。

【０２４５】図２８に記憶装置１０ｂ内に保持されているデータキャッシュ管理情報３４ｂを示す。図１０のデータキャッシュ管理情報３４からの変更点は、ボリュームを利用した記憶管理からファイル利用した記憶管理に変更されたため、キャッシュセグメント情報７２０に修正が加えられ、キャッシュセグメント情報７２０ｂに変更されたことである。キャッシュセグメント情報７２０ｂのキャッシュセグメント情報７２０からの変更点は、上述の理由により、ボリューム名５０１とボリューム論理ブロック番号５１２の部分がファイルシステム名１００１とファイルパス名１００２とファイルブロック番号１００３に変更されたことである。

【０２４６】記憶装置１０ｂにおける本実施の形態における第一の実施の形態からの差は、ほとんどがボリューム名５０１をファイルシステム名１００１とファイルパス名１００２に、ボリューム論理ブロック番号５１２をファイルブロック番号１００３に変更することであり、その他の変更点もその差を述べてきた。記憶装置１０ｂにおける処理に関しても基本的にこれまで述べてきた変更点と同じ変更点への対応方法を実施すれば、第一の実

施の形態における処理をほぼそのまま本実施の形態に当てはめることができる。

【０２４７】

【発明の効果】本発明により以下のことが可能となる。第一に、ＤＢＭＳが管理するデータを保持する記憶装置において、ＤＢＭＳ向けのアクセスの最適化が実現される。この記憶装置を用いることにより、既存のＤＢＭＳに対してプログラムの修正無しにＤＢＭＳ稼動システムの性能を向上させることができる。つまり、高性能なＤＢシステムを容易に構築できる。

【０２４８】第二に、ＤＢＭＳが管理するデータを保持する記憶装置において、ＤＢのデータや処理に与えられた処理優先度を考慮するアクセスの最適化が実現される。ＤＢのデータ毎の処理優先度や処理毎の優先度を考慮することにより、特定のＤＢに対する処理性能を保持するＤＢシステムが実現できる。つまり、ＤＢの処理性能に対するＳＬＡ（Ｓｅｒｖｉｃｅ　Ｌｅｖｅｌ　Ａｇｒｅｅｍｅｎｔ）を守るシステムを容易に実現できる。また、特定のＤＢの処理性能が保持されるシステムを構築できるため、ＤＢシステムの性能に関する管理コストを削減できる。

【図面の簡単な説明】

【図１】第一の実施の形態における計算機システムの構成を示す図である。

【図２】ＤＢホスト８０ａ，８０ｂのＯＳ１００内に記憶されているマッピング情報１０６を示す図である。

【図３】ＤＢＭＳ１１０ａ，１１０ｂ内に記憶されているその内部で定義・管理しているデータその他の管理情報であるスキーマ情報１１４を示す図である。

【図４】記憶装置１０内に保持されているボリューム物理記憶位置管理情報４２を示す図である。

【図５】記憶装置１０内に保持されている物理記憶装置稼動情報３２を示す図である。

【図６】記憶装置１０内に保持されているＤＢＭＳデータ情報４０を示す図である。

【図７】ＤＢＭＳデータ情報４０中に含まれるＤＢＭＳスキーマ情報７１１を示す図である。

【図８】ＤＢＭＳデータ情報４０中に含まれるデータ構造物理記憶位置情報７１２を示す図である。

【図９】記憶装置１０内に保持されているＤＢＭＳ実行情報３８を示す図である。

【図１０】記憶装置１０内に保持されているデータキャッシュ管理情報３４を示す図である。

【図１１】記憶装置１０内に保持されている処理優先度付ディスクＩ／Ｏ管理情報３６を示す図である。

【図１２】クエリ８７１とその処理を実現するためにＤＢＭＳ１１０ａが作成したクエリ実行プラン８７２を図示したものである。

【図１３】記憶装置１０に与えられるクエリ実行プランに関する情報であるクエリプラン情報８８０を示す図である。

【図１４】記憶装置１０がクエリプラン情報８８０を受け取った際の処理フローを示す図である。

【図１５】記憶装置１０がクエリプラン情報８８０に対応するクエリの完了通知を受け取った時の処理フローを示す図である。

【図１６】記憶装置１０がＤＢホスト８０ａ，８０ｂから書き込みアクセス要求を受け取ったときの処理フローを示す図である。

【図１７】記憶装置１０がＤＢホスト８０ａ，８０ｂから読み出しアクセス要求を受け取ったときの処理フローを示す図である。

【図１８】ＤＢデータの読み出しアクセス後の処理フローを示す図（その１）である。

【図１９】ＤＢデータの読み出しアクセス後の処理フローを示す図（その２）である。

【図２０】ディスクＩ／Ｏ実行管理情報８６０を利用した物理記憶装置１８へのアクセス処理を行うバックグラウンド処理の処理フローを示す図である。

【図２１】バックグラウンドで実行するダーティセグメント数と再利用ＬＲＵリスト中存在セグメント数の管理のための周期処理の処理フローを示す図である。

【図２２】第二の実施の形態における計算機システムの構成を示す図である。

【図２３】ＤＢホスト８０ｃ，８０ｄのＯＳ１００内に記憶されているネットワークマウント情報１０６ｂを示す図である。

【図２４】記憶装置１０ｂ内に保持されるファイル記憶管理情報４２ｂを示す図である。

【図２５】記憶装置１０ｂ内に保持される物理記憶装置稼動情報３２ｂを示す図である。

【図２６】記憶装置１０ｂ内に保持されているＤＢＭＳデータ情報４０ｂを示す図である。

【図２７】ＤＢＭＳデータ情報４０ｂ中に含まれるデータ構造物理記憶位置情報７１２ｂを示す図である。

【図２８】記憶装置１０ｂ内に保持されているデータキャッシュ管理情報３４ｂを示す図である。

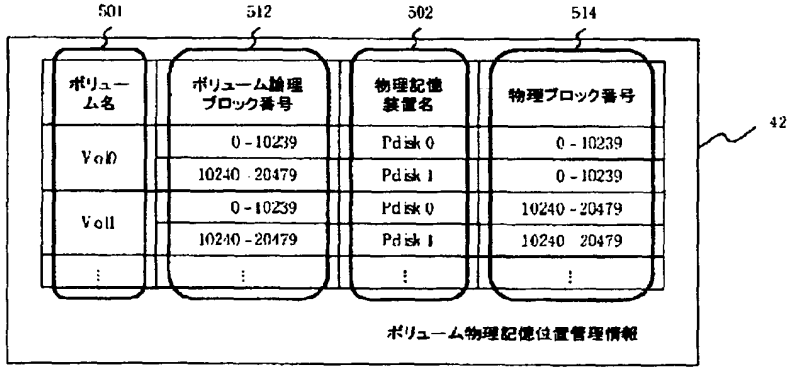
【符号の説明】

| | |
|---|---|
| １０，１０ｂ | 記憶装置 |
| １８ | 物理記憶装置 |
| ２８ | データキャッシュ |
| ３２，３２ｂ | 物理記憶装置稼動情報 |
| ３４，３４ｂ | データキャッシュ管理情報 |
| ３６，３６ｂ | 処理優先度付ディスクＩ／Ｏ管理情報 |
| ３８ | ＤＢＭＳ実行情報 |
| ４０，４０ｂ | ＤＢＭＳデータ情報 |
| ４２ | ボリューム物理記憶位置管理情報 |

| | | | |
|---|---|---|---|
| ４２ｂ | ファイル記憶管理情報 | １００ | ＯＳ（オペレーティングシステム） |
| ５０ | 記憶装置制御プログラム | １０２ | ボリュームマネージャ |
| ５２ | ディスクコントローラ制御部 | １０４ | ファイルシステム |
| ５４ | キャッシュ制御部 | １０４ｂ | ネットワークファイルシステム |
| ５６ | 物理記憶位置管理部 | １０６ | マッピング情報 |
| ５８ | Ｉ／Ｏパスインターフェイス制御部 | １０６ｂ | ネットワークマウント情報 |
| ６０ | ネットワークインターフェイス制御部 | １１０ａ，１１０ｂ | ＤＢＭＳ（データベース管理システム） |
| ７０ | Ｉ／Ｏパスインターフェイス | １１４ | スキーマ情報 |
| ７１ | Ｉ／Ｏパス | １１６ | ＤＢＭＳ情報通信部 |
| ７２ | Ｉ／Ｏパススイッチ | １１８ | ＤＢＭＳ情報取得・通信プログラム |
| ７８ | ネットワークインターフェイス | １２０ | クエリプラン取得プログラム |
| ７９ | ネットワーク | １２６ | ＤＢＭＳフロントエンドプログラム |
| ８０ａ，８０ｂ，８０ｃ，８０ｄ | ＤＢホスト | １３０ | ホスト情報設定プログラム |
| ８１ | ＤＢクライアント | １３２ | 処理性能管理プログラム |
| ８２ | 処理性能管理サーバ | | |

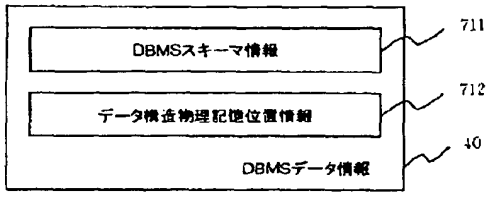
【図４】

図4



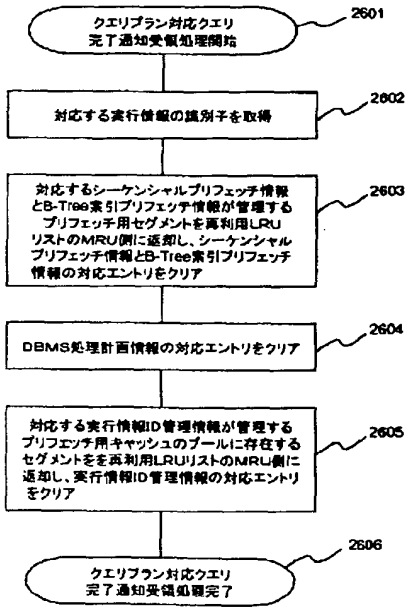

【図６】

図6



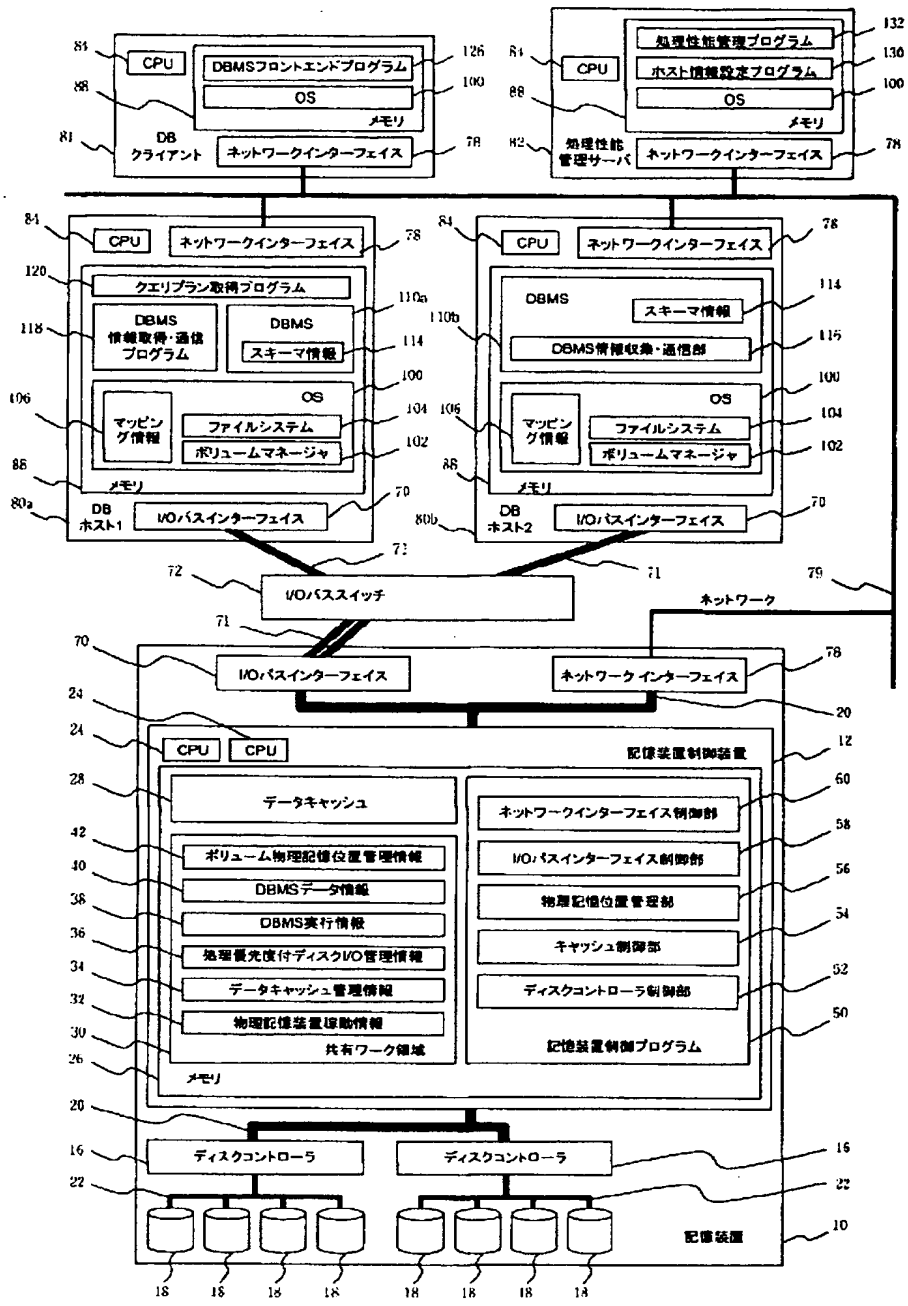

【図１５】

図15



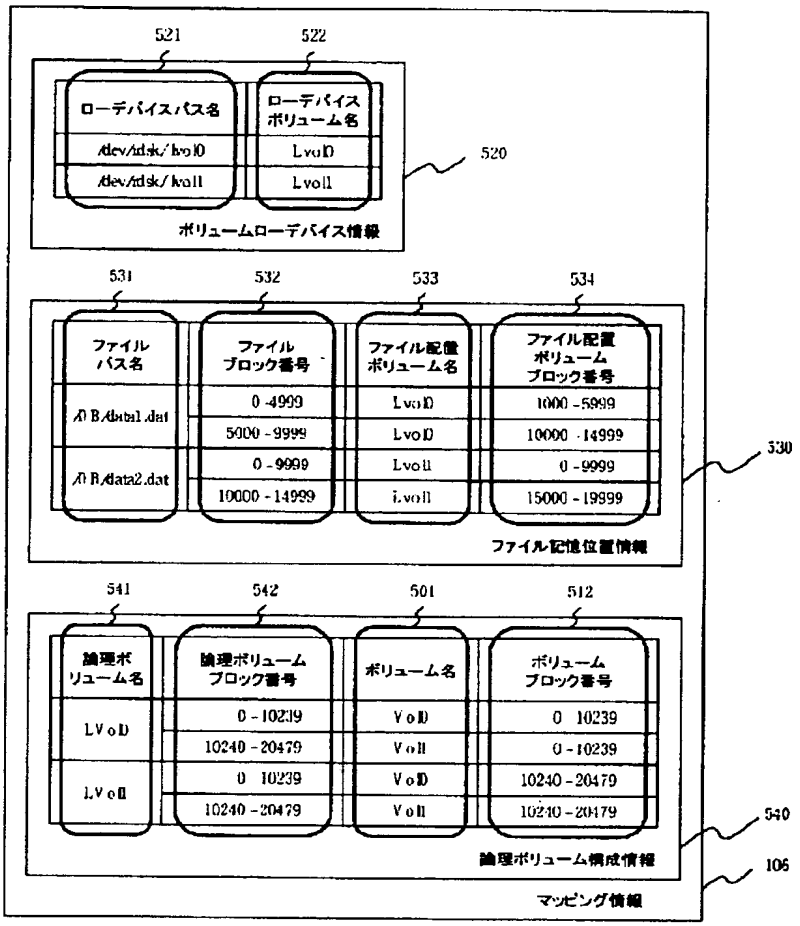
【図１】

図1


CPU
DBMSフロントエンドプログラム
OS
メモリ
DB
クライアント
ネットワークインターフェイス
処理性能管理プログラム
ホスト情報設定プログラム
処理性能
管理サーバ
クエリプラン取得プログラム
DBMS
情報取得・通信
プログラム
スキーマ情報
DBMS情報収集・通信部
マッピング情報
ファイルシステム
ボリュームマネージャ
DB
ホスト1
DB
ホスト2
I/Oパススイッチ
ネットワーク
I/Oパスインターフェイス
ネットワーク インターフェイス
記憶装置制御装置
データキャッシュ
ボリューム物理記憶位置管理情報
DBMSデータ情報
DBMS実行情報
処理優先度付ディスクI/O管理情報
データキャッシュ管理情報
物理記憶装置増設情報
共有ワーク領域
メモリ
ネットワークインターフェイス制御部
I/Oパスインターフェイス制御部
物理記憶位置管理部
キャッシュ制御部
ディスクコントローラ制御部
記憶装置制御プログラム
ディスクコントローラ
記憶装置

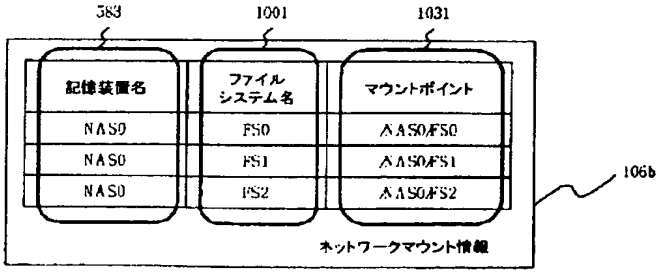
【図２】

図2


521
522
ローデバイスパス名
ローデバイスボリューム名
/dev/rdsk/lvol0
Lvol0
/dev/rdsk/lvol1
Lvol1
520
ボリュームローデバイス情報
531
532
533
534
ファイルパス名
ファイルブロック番号
ファイル配置ボリューム名
ファイル配置ボリュームブロック番号
/0 B/data1.dat
0 -4999
Lvol0
1000 -5999
5000 - 9999
Lvol0
10000 -14999
/0 B/data2.dat
0 - 9999
Lvol1
0 - 9999
10000 - 14999
Lvol1
15000 - 19999
530
ファイル記憶位置情報
541
542
501
512
論理ボリューム名
論理ボリュームブロック番号
ボリューム名
ボリュームブロック番号
LVol0
0 - 10239
Vol0
0 10239
10240 - 20479
Vol1
0 - 10239
LVol1
0 10239
Vol0
10240 - 20479
10240 - 20479
Vol1
10240 - 20479
540
論理ボリューム構成情報
106
マッピング情報

【図２３】

図23


583
1001
1031
記憶装置名
ファイルシステム名
マウントポイント
NAS0
FS0
/NAS0/FS0
NAS0
FS1
/NAS0/FS1
NAS0
FS2
/NAS0/FS2
106b
ネットワークマウント情報

【図３】

図3

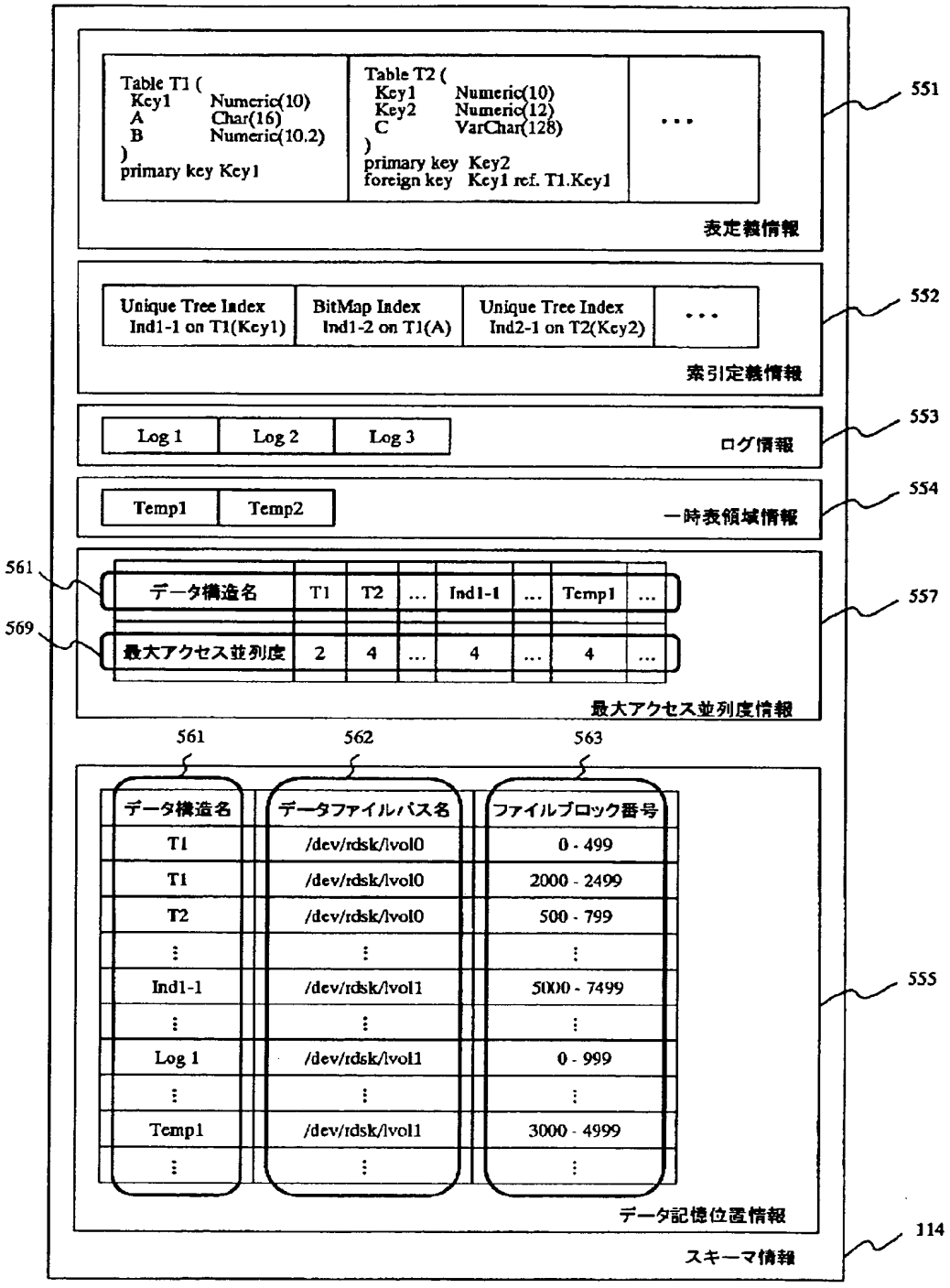
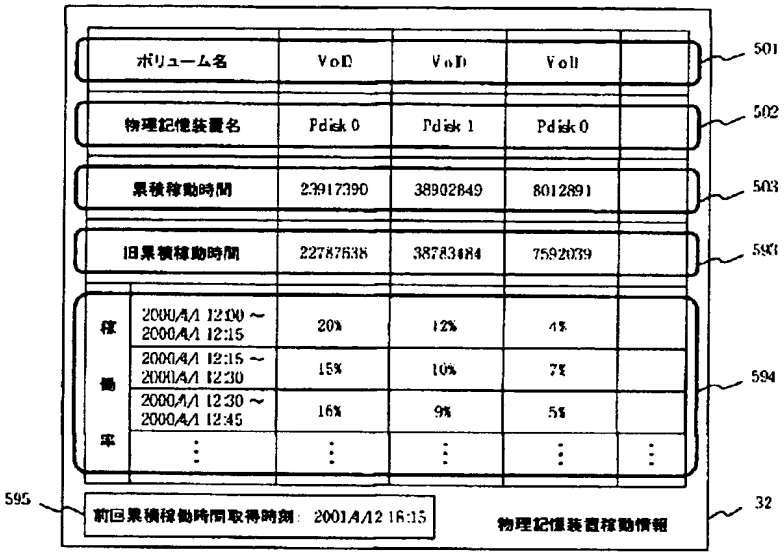


Table T1 (
Key1 Numeric(10)
A Char(16)
B Numeric(10.2)
)
primary key Key1
Table T2 (
Key1 Numeric(10)
Key2 Numeric(12)
C VarChar(128)
)
primary key Key2
foreign key Key1 ref. T1.Key1
...
551
表定義情報
Unique Tree Index Ind1-1 on T1(Key1)
BitMap Index Ind1-2 on T1(A)
Unique Tree Index Ind2-1 on T2(Key2)
...
552
索引定義情報
Log 1
Log 2
Log 3
553
ログ情報
Temp1
Temp2
554
一時表領域情報
561
569
データ構造名
T1
T2
...
Ind1-1
...
Temp1
...
最大アクセス並列度
2
4
...
4
...
4
...
557
最大アクセス並列度情報
561
562
563
データ構造名
データファイルパス名
ファイルブロック番号
T1
/dev/rdsk/lvol0
0 - 499
T1
/dev/rdsk/lvol0
2000 - 2499
T2
/dev/rdsk/lvol0
500 - 799
Ind1-1
/dev/rdsk/lvol1
5000 - 7499
Log 1
/dev/rdsk/lvol1
0 - 999
Temp1
/dev/rdsk/lvol1
3000 - 4999
555
データ記憶位置情報
114
スキーマ情報

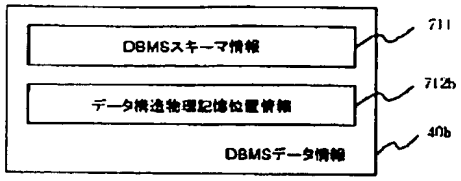
【図５】

図5



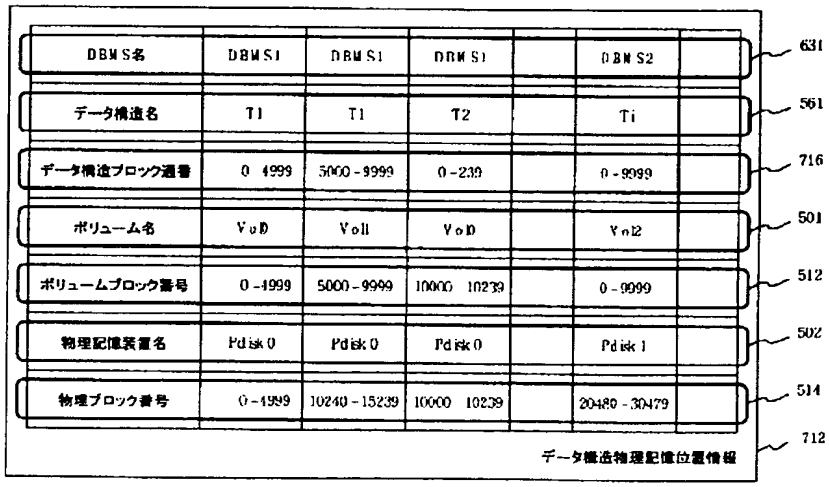

【図２６】

図26



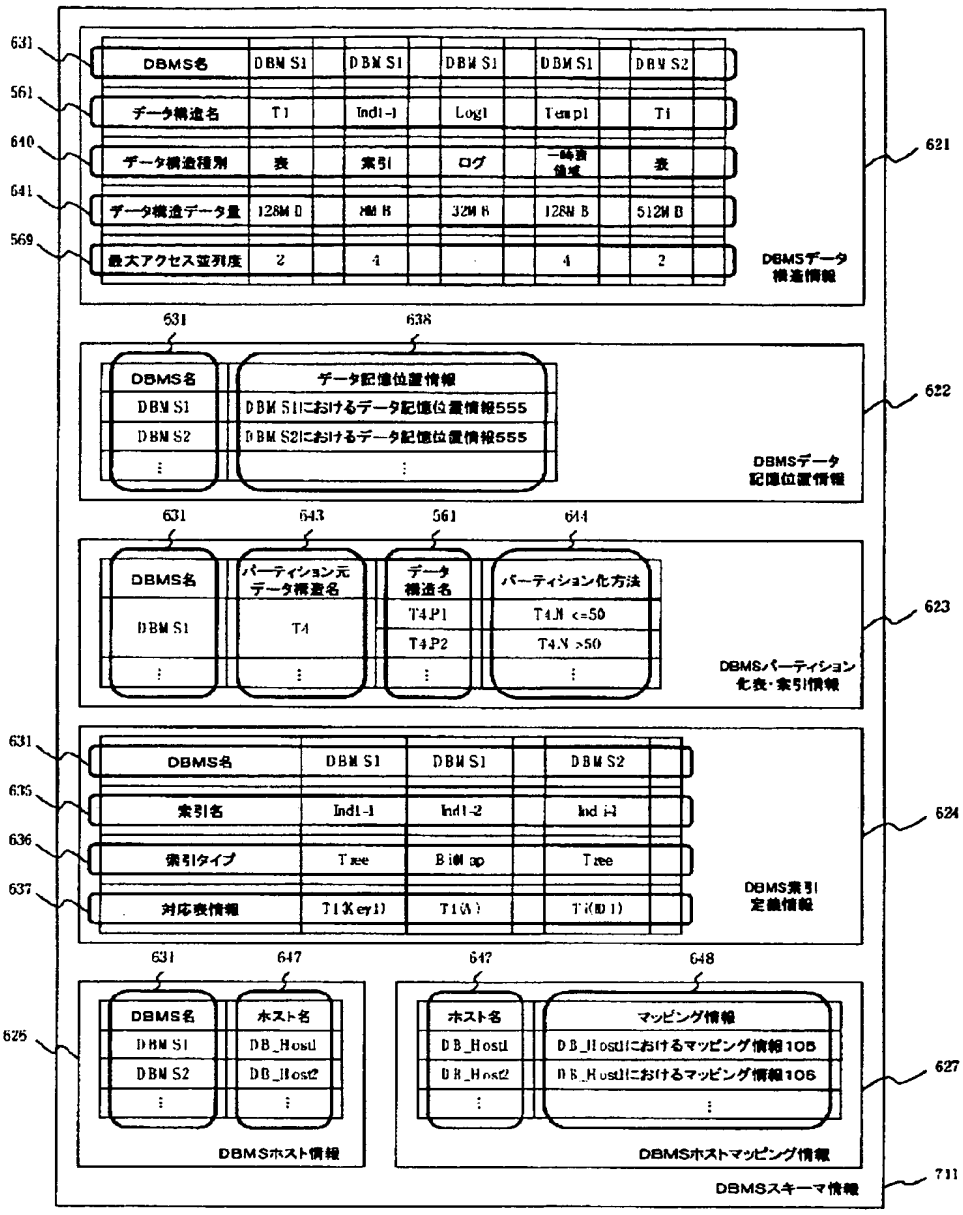

【図８】

図8



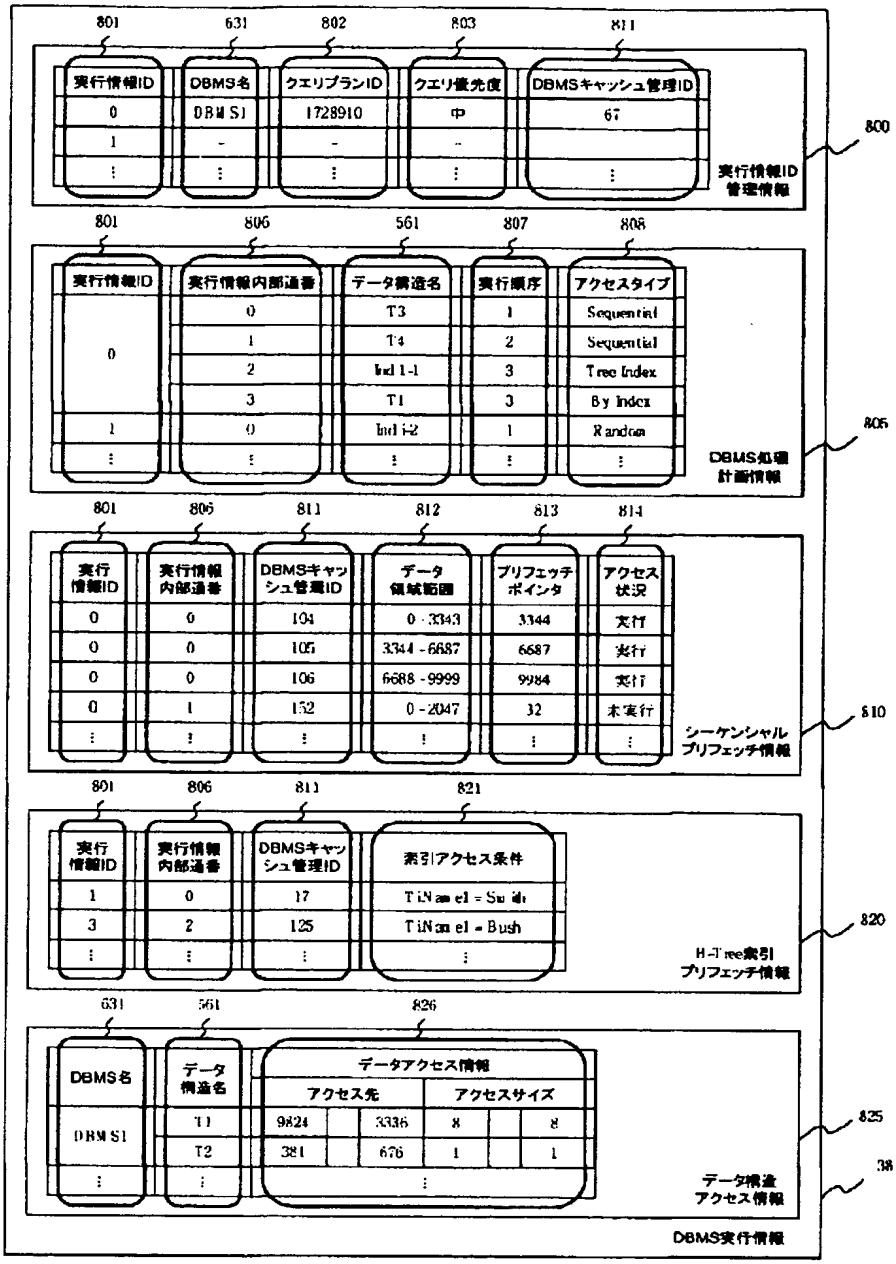
【図７】

図7


631
561
640
641
569
DBMS名
DBMS1
DBMS1
DBMS1
DBMS1
DBMS2
データ構造名
T1
Ind1-1
Log1
Temp1
T1
データ構造種別
表
索引
ログ
一時表領域
表
データ構造データ量
128MB
MMB
32MB
128MB
512MB
最大アクセス並列度
2
4
4
2
621
DBMSデータ構造情報
631
638
DBMS名
データ記憶位置情報
DBMS1
DBMS1におけるデータ記憶位置情報555
DBMS2
DBMS2におけるデータ記憶位置情報555
622
DBMSデータ記憶位置情報
631
643
561
644
DBMS名
パーティション元データ構造名
データ構造名
パーティション化方法
DBMS1
T4
T4.P1
T4.N <=50
T4.P2
T4.N >50
623
DBMSパーティション化表・索引情報
631
635
636
637
DBMS名
DBMS1
DBMS1
DBMS2
索引名
Ind1-1
Ind1-2
Ind 1-3
索引タイプ
Tree
BitMap
Tree
対応表情報
T1(Key1)
T1(A)
T1(B1)
624
DBMS索引定義情報
631
647
647
648
625
DBMS名
ホスト名
DBMS1
DB_Host1
DBMS2
DB_Host2
DBMSホスト情報
ホスト名
マッピング情報
DB_Host1
DB_Host1におけるマッピング情報105
DB_Host2
DB_Host1におけるマッピング情報106
627
DBMSホストマッピング情報
711
DBMSスキーマ情報

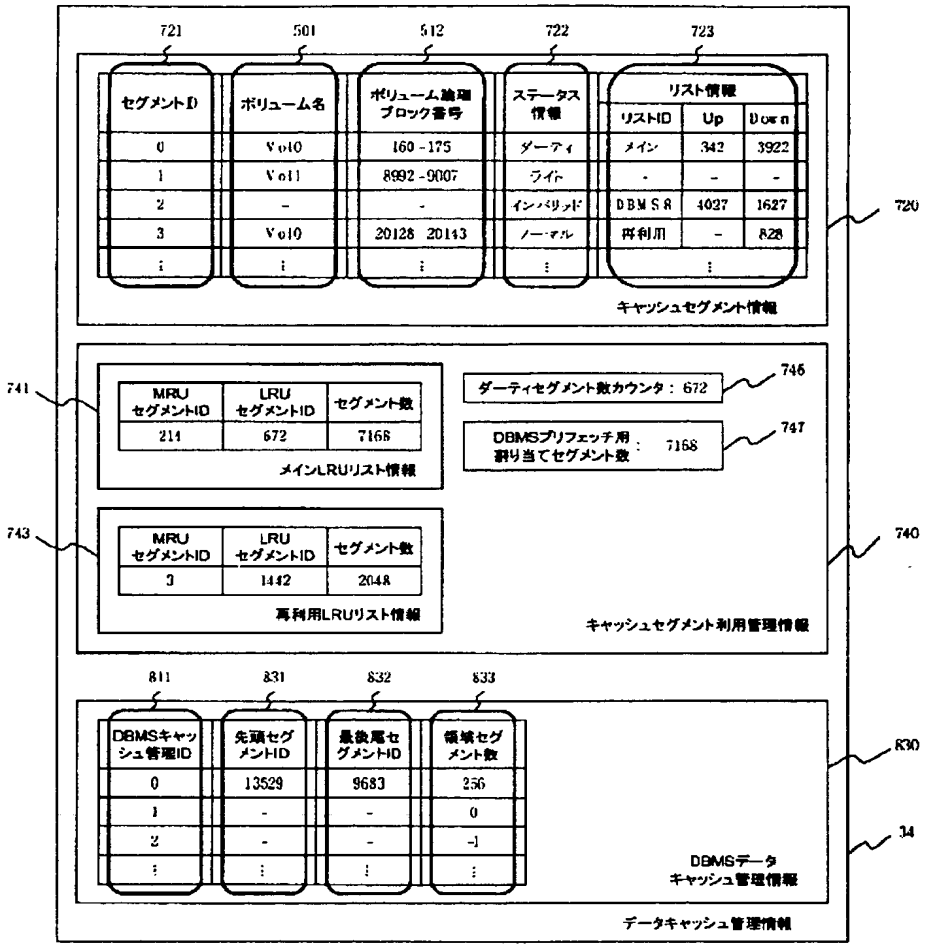
【図９】

図9


801
631
802
803
811
実行情報ID
DBMS名
クエリプランID
クエリ優先度
DBMSキャッシュ管理ID
0
DBMS1
1728910
中
67
1
-
-
-
800
実行情報ID管理情報
801
806
561
807
808
実行情報ID
実行情報内部通番
データ構造名
実行順序
アクセスタイプ
0
0
T3
1
Sequential
1
T4
2
Sequential
2
Ind 1-1
3
Tree Index
3
T1
3
By Index
1
0
Ind 1-2
1
Random
805
DBMS処理計画情報
801
806
811
812
813
814
実行情報ID
実行情報内部通番
DBMSキャッシュ管理ID
データ領域範囲
プリフェッチポインタ
アクセス状況
0
0
104
0 - 3343
3344
実行
0
0
105
3344 - 6687
6687
実行
0
0
106
6688 - 9999
9984
実行
0
1
152
0 - 2047
32
未実行
810
シーケンシャルプリフェッチ情報
801
806
811
821
実行情報ID
実行情報内部通番
DBMSキャッシュ管理ID
索引アクセス条件
1
0
17
TiName1 = Smith
3
2
125
TiName1 = Bush
820
H-Tree索引プリフェッチ情報
631
561
826
DBMS名
データ構造名
データアクセス情報
アクセス先
アクセスサイズ
DBMS1
T1
9824
3336
8
8
T2
381
676
1
1
825
データ構造アクセス情報
38
DBMS実行情報

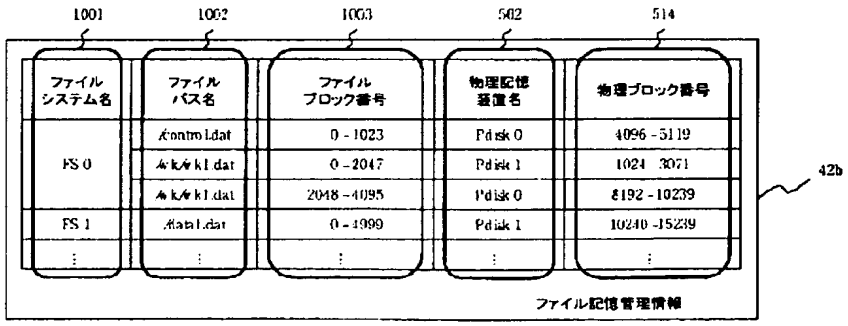
【図１０】

図10



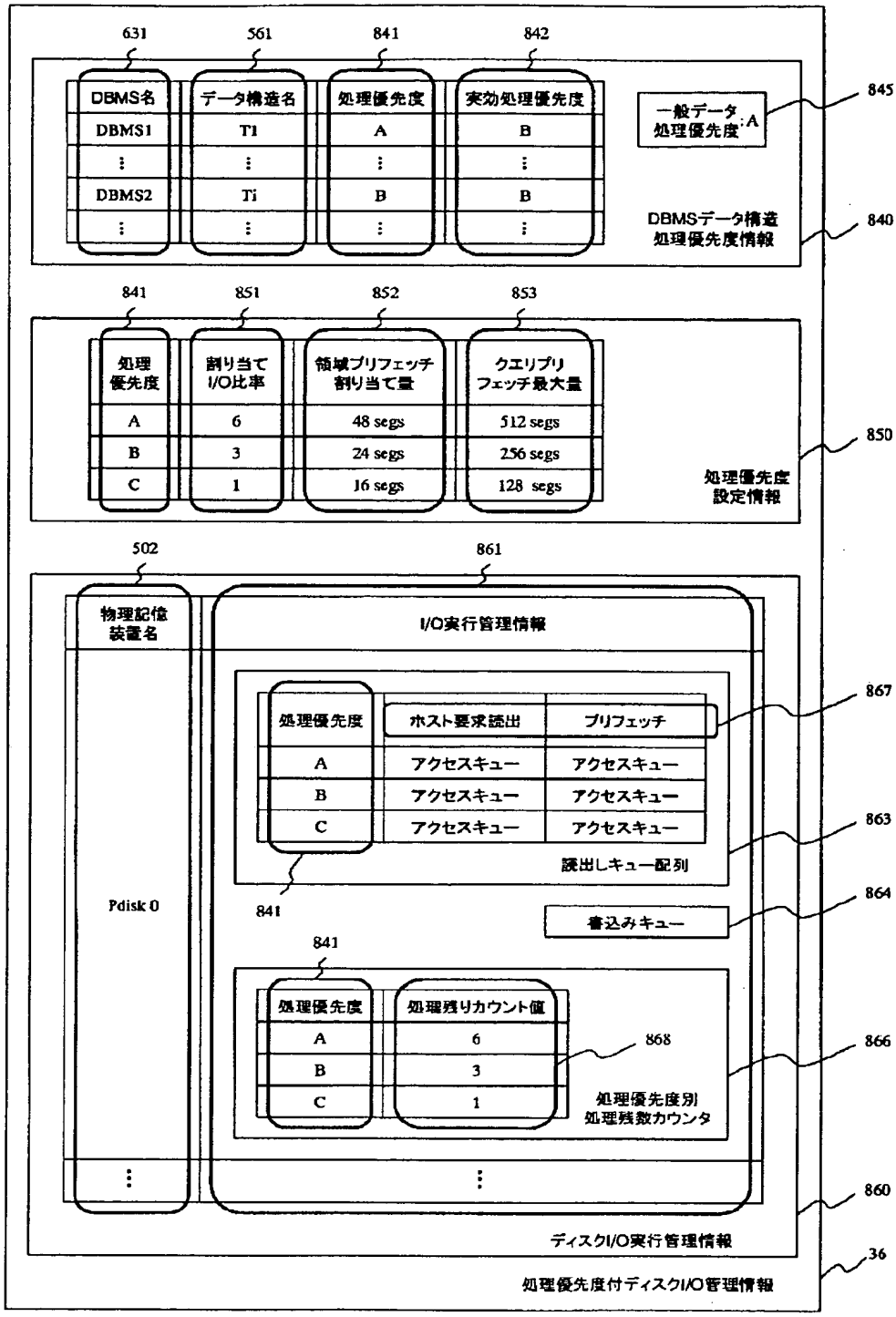

【図２４】

図24



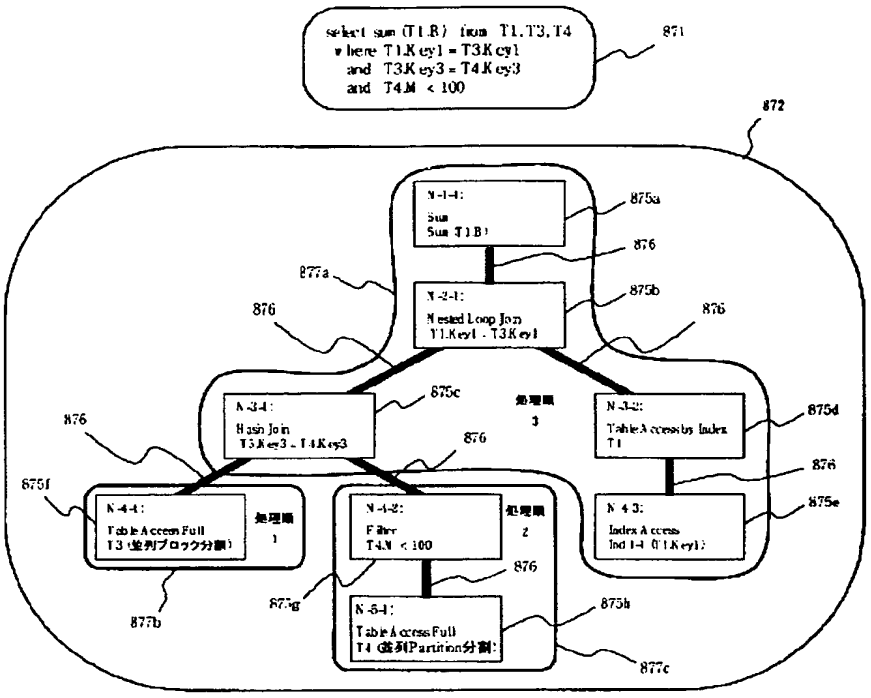
【図１１】

図11


631
561
841
842
DBMS名
データ構造名
処理優先度
実効処理優先度
DBMS1
T1
A
B
DBMS2
Ti
B
B
一般データ処理優先度:A
845
DBMSデータ構造処理優先度情報
840
841
851
852
853
処理優先度
割り当てI/O比率
領域プリフェッチ割り当て量
クエリプリフェッチ最大量
A
6
48 segs
512 segs
B
3
24 segs
256 segs
C
1
16 segs
128 segs
850
処理優先度設定情報
502
861
物理記憶装置名
I/O実行管理情報
処理優先度
ホスト要求読出
プリフェッチ
A
アクセスキュー
アクセスキュー
B
アクセスキュー
アクセスキュー
C
アクセスキュー
アクセスキュー
867
863
読出しキュー配列
841
Pdisk 0
書込みキュー
864
841
処理優先度
処理残りカウント値
A
6
B
3
C
1
868
866
処理優先度別処理残数カウンタ
860
ディスクI/O実行管理情報
36
処理優先度付ディスクI/O管理情報

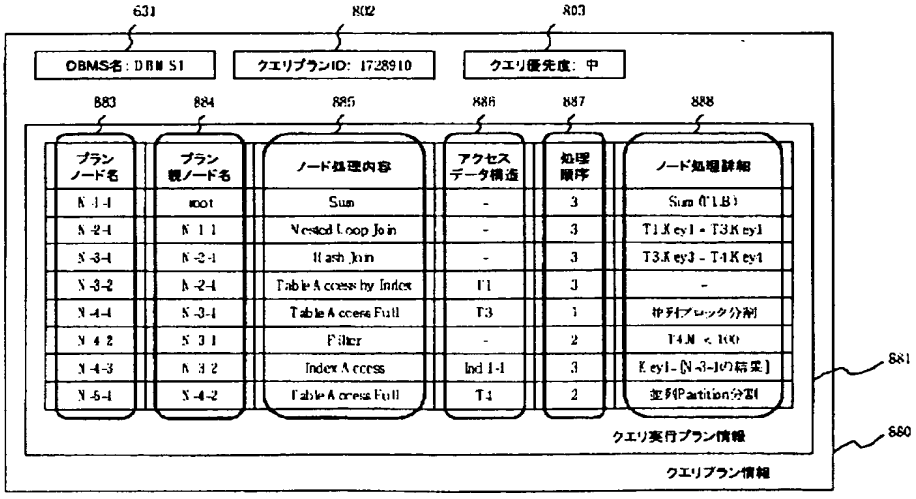
【図１２】

図12



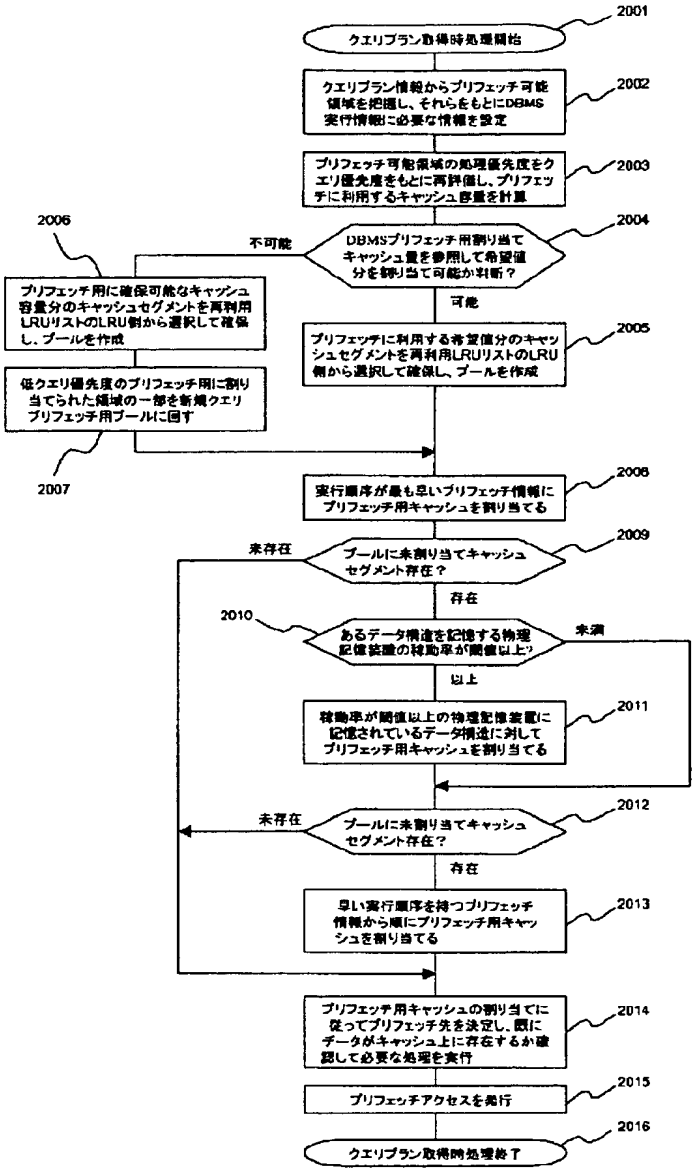

【図１３】

図13



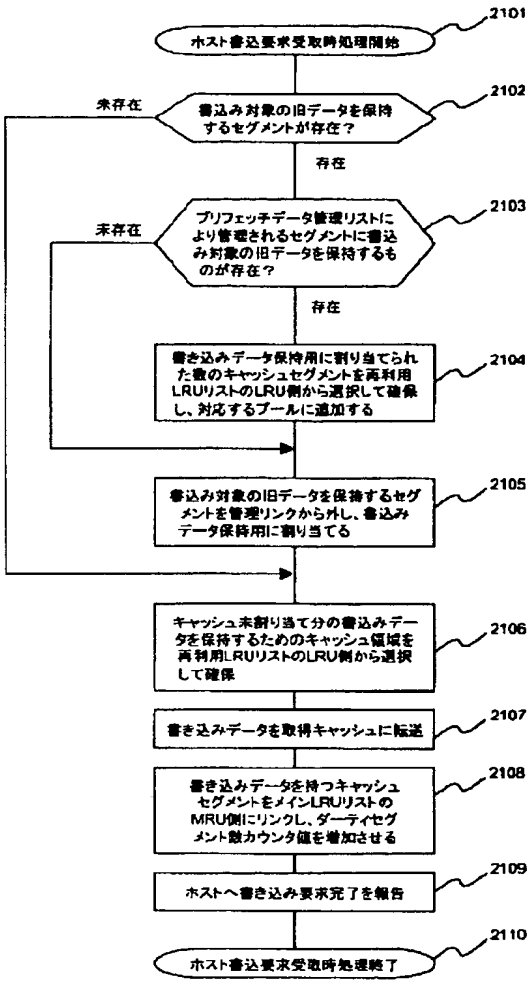
【図１４】

図14



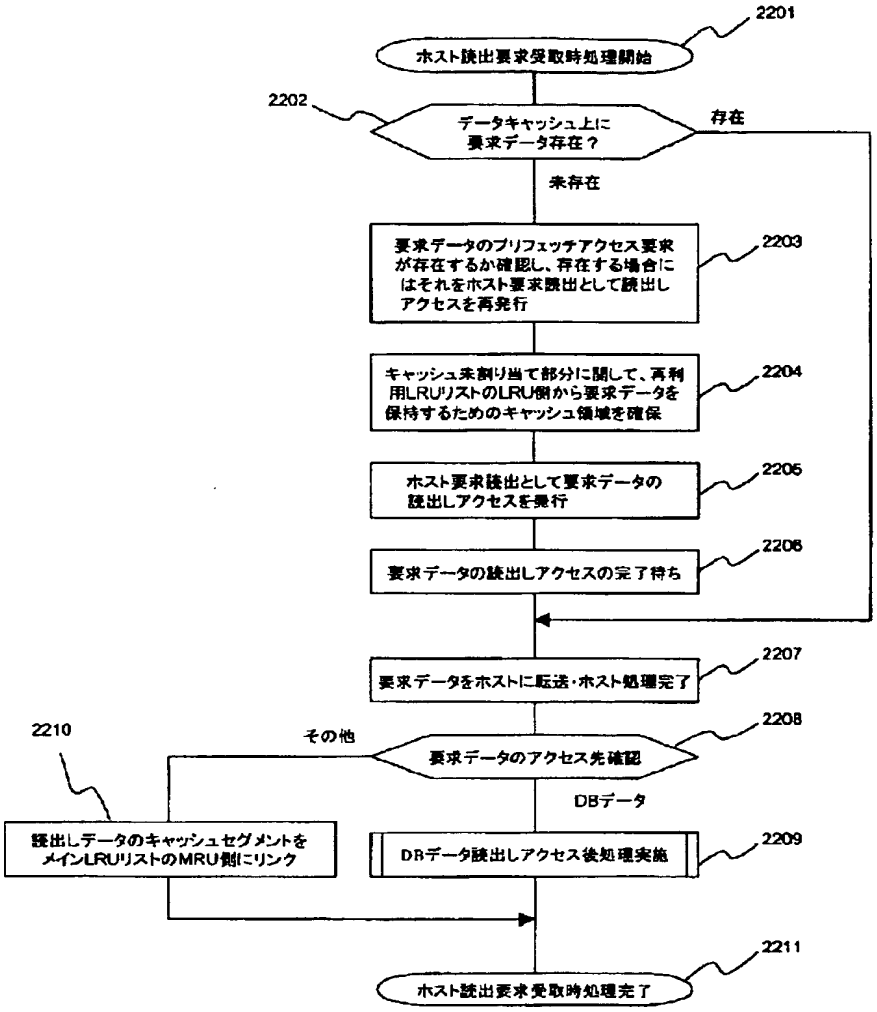

【図１６】

図16



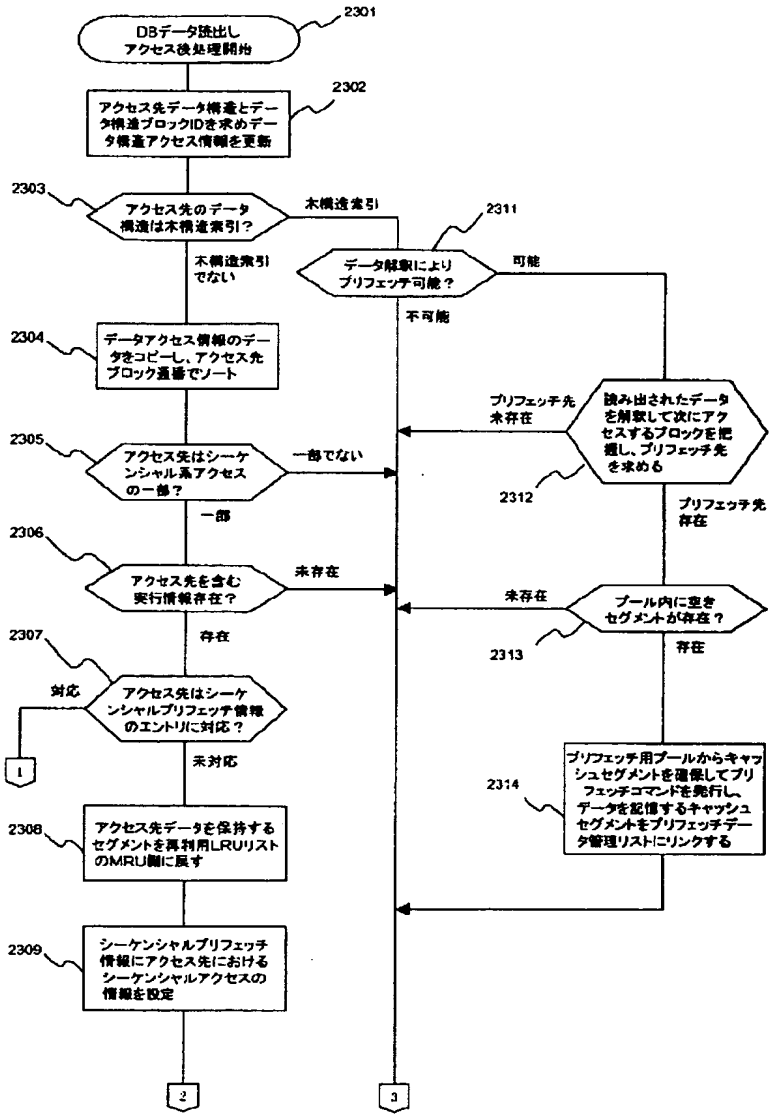
【図１７】

図17


2201
ホスト読出要求受取時処理開始
2202
データキャッシュ上に要求データ存在？
存在
未存在
要求データのプリフェッチアクセス要求が存在するか確認し、存在する場合にはそれをホスト要求読出として読出しアクセスを再発行
2203
キャッシュ未割り当て部分に関して、再利用LRUリストのLRU側から要求データを保持するためのキャッシュ領域を確保
2204
ホスト要求読出として要求データの読出しアクセスを発行
2205
要求データの読出しアクセスの完了待ち
2206
要求データをホストに転送・ホスト処理完了
2207
要求データのアクセス先確認
2208
その他
DBデータ
DBデータ読出しアクセス後処理実施
2209
読出しデータのキャッシュセグメントをメインLRUリストのMRU側にリンク
2210
ホスト読出要求受取時処理完了
2211

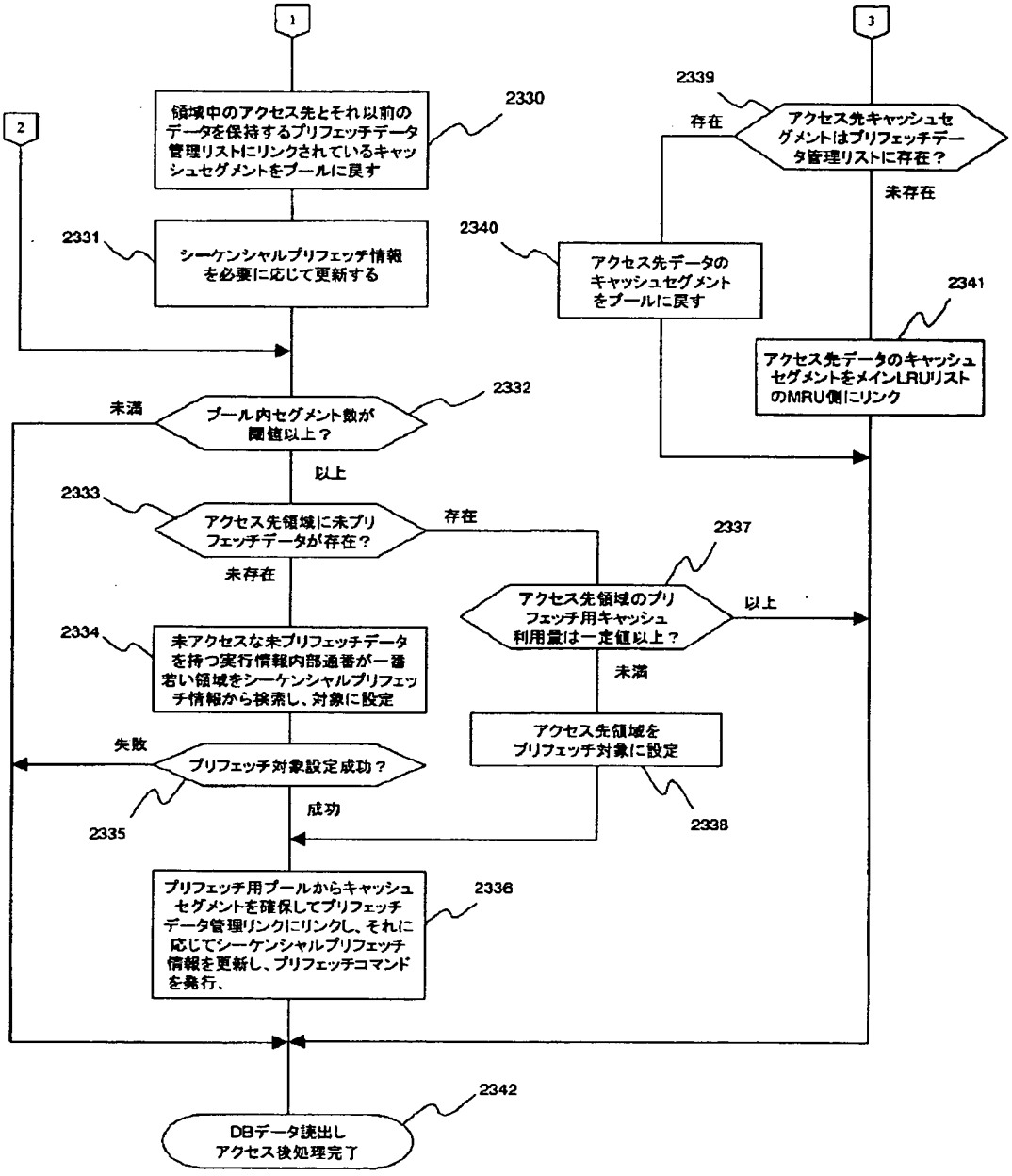
【図１８】

図18


DBデータ読出し
アクセス後処理開始
2301
アクセス先データ構造とデータ構造ブロックIDを求めデータ構造アクセス情報を更新
2302
2303
アクセス先のデータ構造は木構造索引?
木構造索引
木構造索引でない
2311
データ解釈によりプリフェッチ可能?
可能
不可能
2304
データアクセス情報のデータをコピーし、アクセス先ブロック番号でソート
プリフェッチ先未存在
読み出されたデータを解釈して次にアクセスするブロックを把握し、プリフェッチ先を求める
2312
プリフェッチ先存在
2305
アクセス先はシーケンシャル系アクセスの一部?
一部でない
一部
2306
アクセス先を含む実行情報存在?
未存在
未存在
プール内に空きセグメントが存在?
2313
存在
存在
2307
アクセス先はシーケンシャルプリフェッチ情報のエントリに対応?
対応
未対応
1
2314
プリフェッチ用プールからキャッシュセグメントを確保してプリフェッチコマンドを発行し、データを記憶するキャッシュセグメントをプリフェッチデータ管理リストにリンクする
2308
アクセス先データを保持するセグメントを再利用LRUリストのMRU側に戻す
2309
シーケンシャルプリフェッチ情報にアクセス先におけるシーケンシャルアクセスの情報を設定
2
3

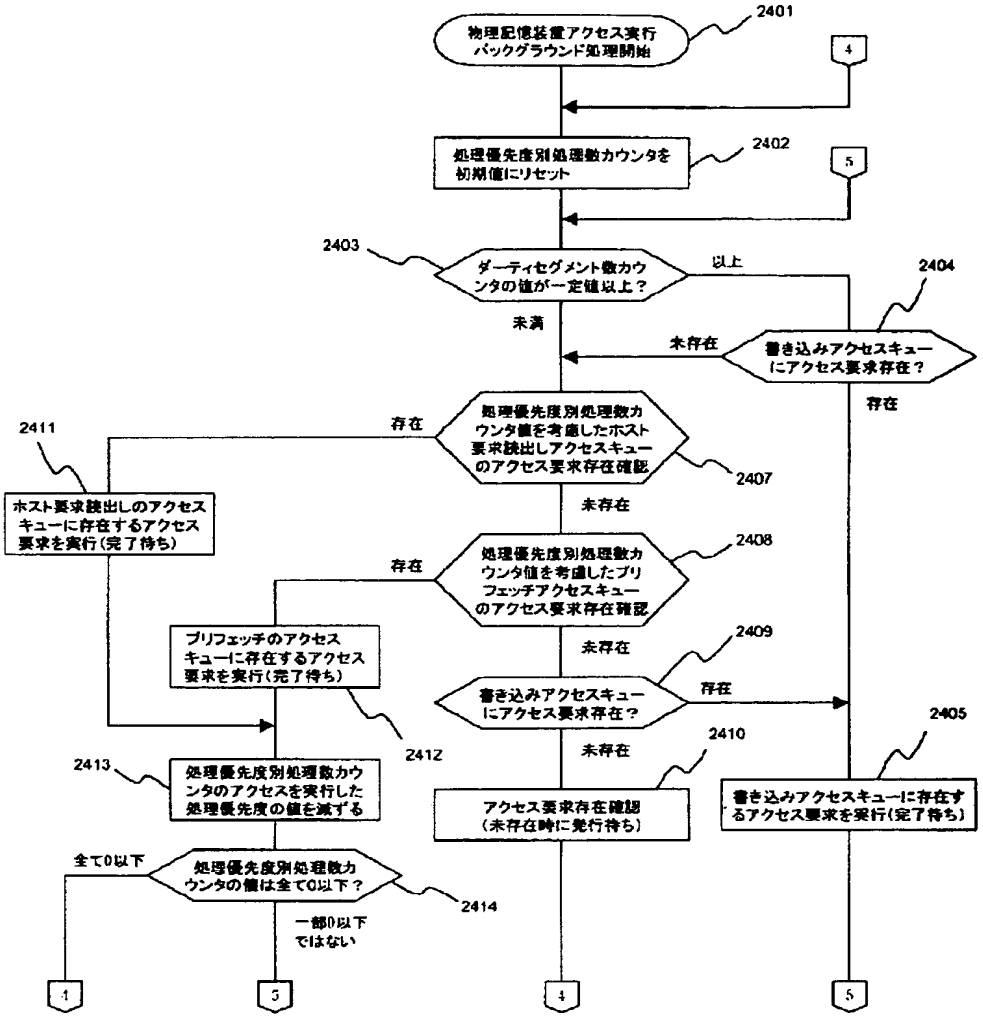
【図１９】

図19


1
2
3
領域中のアクセス先とそれ以前のデータを保持するプリフェッチデータ管理リストにリンクされているキャッシュセグメントをプールに戻す
2330
シーケンシャルプリフェッチ情報を必要に応じて更新する
2331
プール内セグメント数が閾値以上？
2332
未満
以上
アクセス先領域に未プリフェッチデータが存在？
2333
存在
未存在
未アクセスな未プリフェッチデータを持つ実行情報内部通番が一番若い領域をシーケンシャルプリフェッチ情報から検索し、対象に設定
2334
プリフェッチ対象設定成功？
失敗
成功
2335
プリフェッチ用プールからキャッシュセグメントを確保してプリフェッチデータ管理リンクにリンクし、それに応じてシーケンシャルプリフェッチ情報を更新し、プリフェッチコマンドを発行、
2336
DBデータ読出し
アクセス後処理完了
2342
アクセス先キャッシュセグメントはプリフェッチデータ管理リストに存在？
2339
存在
未存在
アクセス先データのキャッシュセグメントをプールに戻す
2340
アクセス先データのキャッシュセグメントをメインLRUリストのMRU側にリンク
2341
アクセス先領域のプリフェッチ用キャッシュ利用量は一定値以上？
2337
以上
未満
アクセス先領域をプリフェッチ対象に設定
2338

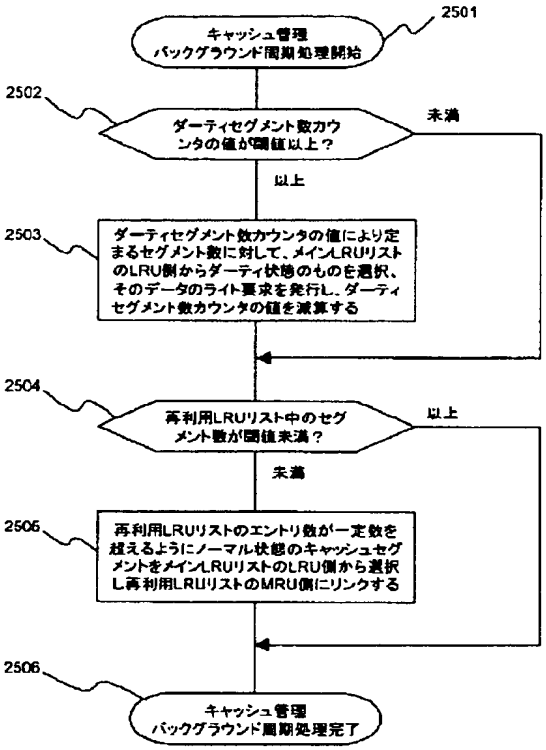
【図２０】

図20


物理記憶装置アクセス実行バックグラウンド処理開始
2401
4
処理優先度別処理数カウンタを初期値にリセット
2402
5
ダーティセグメント数カウンタの値が一定値以上？
2403
以上
2404
未満
未存在
書き込みアクセスキューにアクセス要求存在？
存在
処理優先度別処理数カウンタ値を考慮したホスト要求読出しアクセスキューのアクセス要求存在確認
2407
2411
存在
ホスト要求読出しのアクセスキューに存在するアクセス要求を実行(完了待ち)
未存在
処理優先度別処理数カウンタ値を考慮したプリフェッチアクセスキューのアクセス要求存在確認
2408
存在
プリフェッチのアクセスキューに存在するアクセス要求を実行(完了待ち)
2412
未存在
書き込みアクセスキューにアクセス要求存在？
2409
存在
2405
未存在
2410
2413
処理優先度別処理数カウンタのアクセスを実行した処理優先度の値を減ずる
アクセス要求存在確認(未存在時に発行待ち)
書き込みアクセスキューに存在するアクセス要求を実行(完了待ち)
全て0以下
処理優先度別処理数カウンタの値は全て0以下？
2414
一部0以下ではない
1
3
4
5

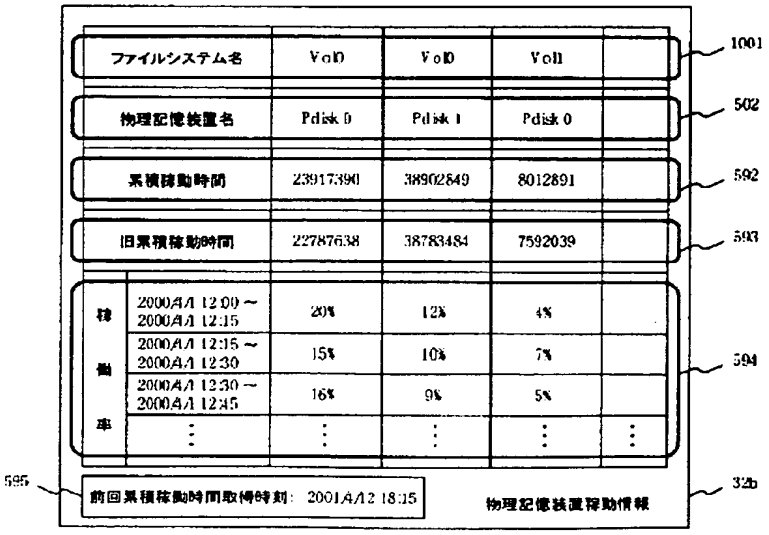
【図２１】

図21



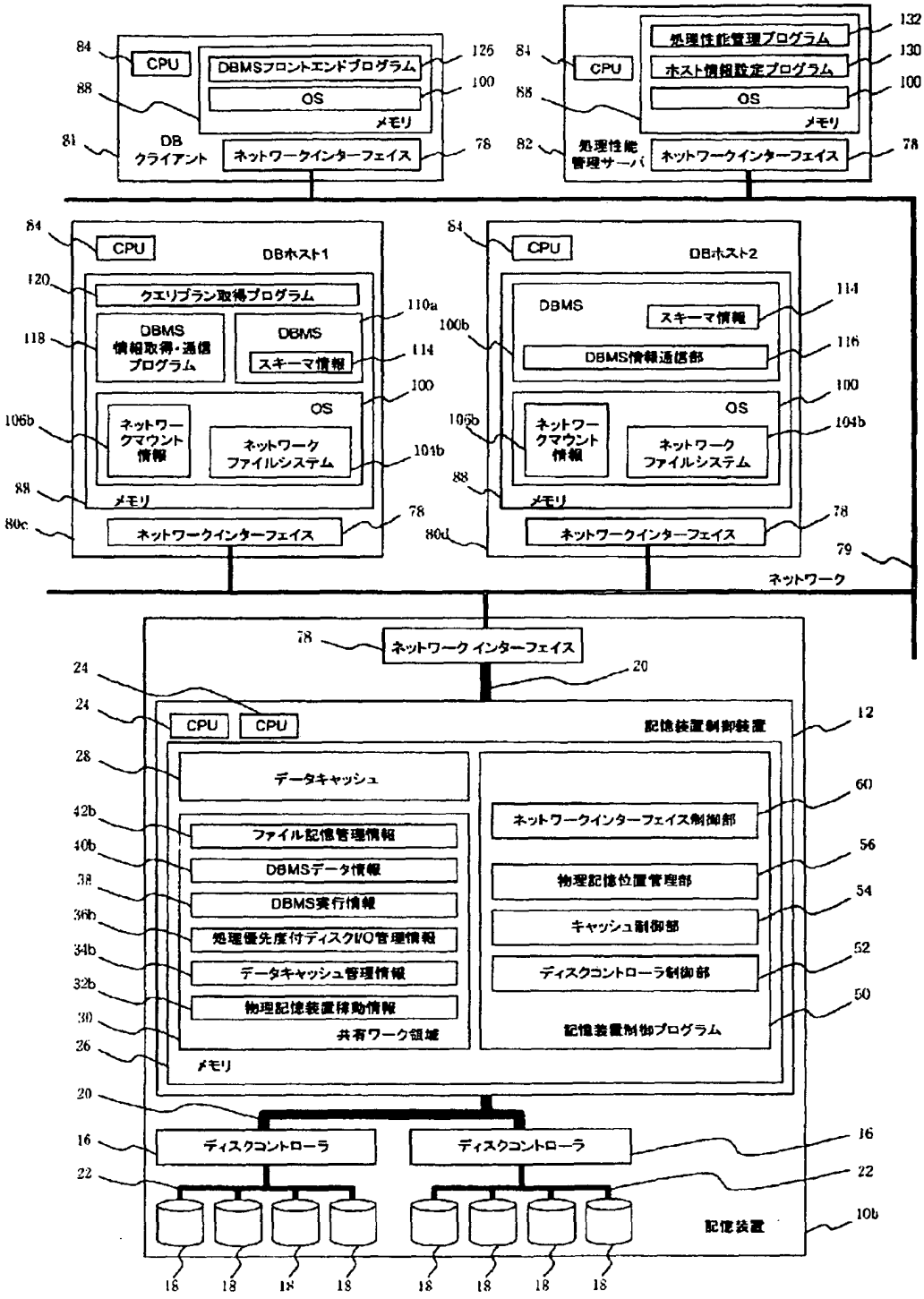

【図２５】

図25



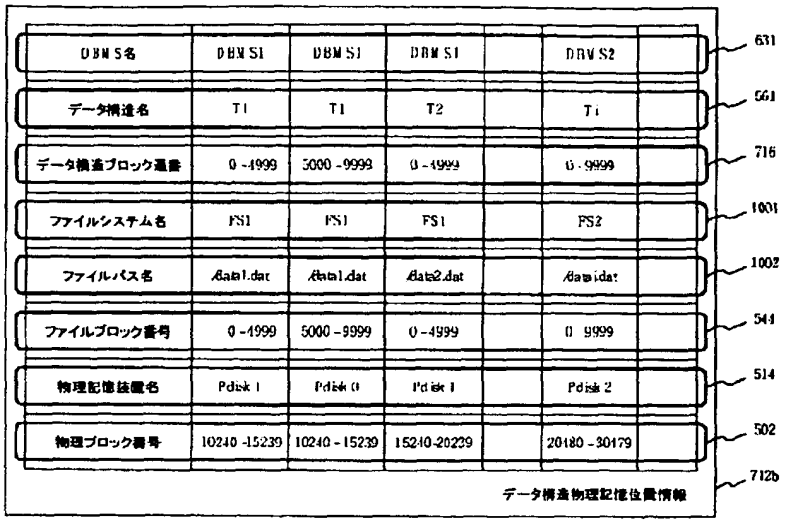
【図２２】

図22


CPU
DBMSフロントエンドプログラム
OS
メモリ
DB
クライアント
ネットワークインターフェイス
処理性能管理プログラム
ホスト情報設定プログラム
処理性能
管理サーバ
DBホスト1
DBホスト2
クエリプラン取得プログラム
DBMS
情報取得・通信
プログラム
スキーマ情報
DBMS情報通信部
ネットワーク
マウント
情報
ネットワーク
ファイルシステム
ネットワーク
ネットワーク インターフェイス
記憶装置制御装置
データキャッシュ
ファイル記憶管理情報
DBMSデータ情報
DBMS実行情報
処理優先度付ディスクI/O管理情報
データキャッシュ管理情報
物理記憶装置稼動情報
共有ワーク領域
メモリ
ネットワークインターフェイス制御部
物理記憶位置管理部
キャッシュ制御部
ディスクコントローラ制御部
記憶装置制御プログラム
ディスクコントローラ
記憶装置

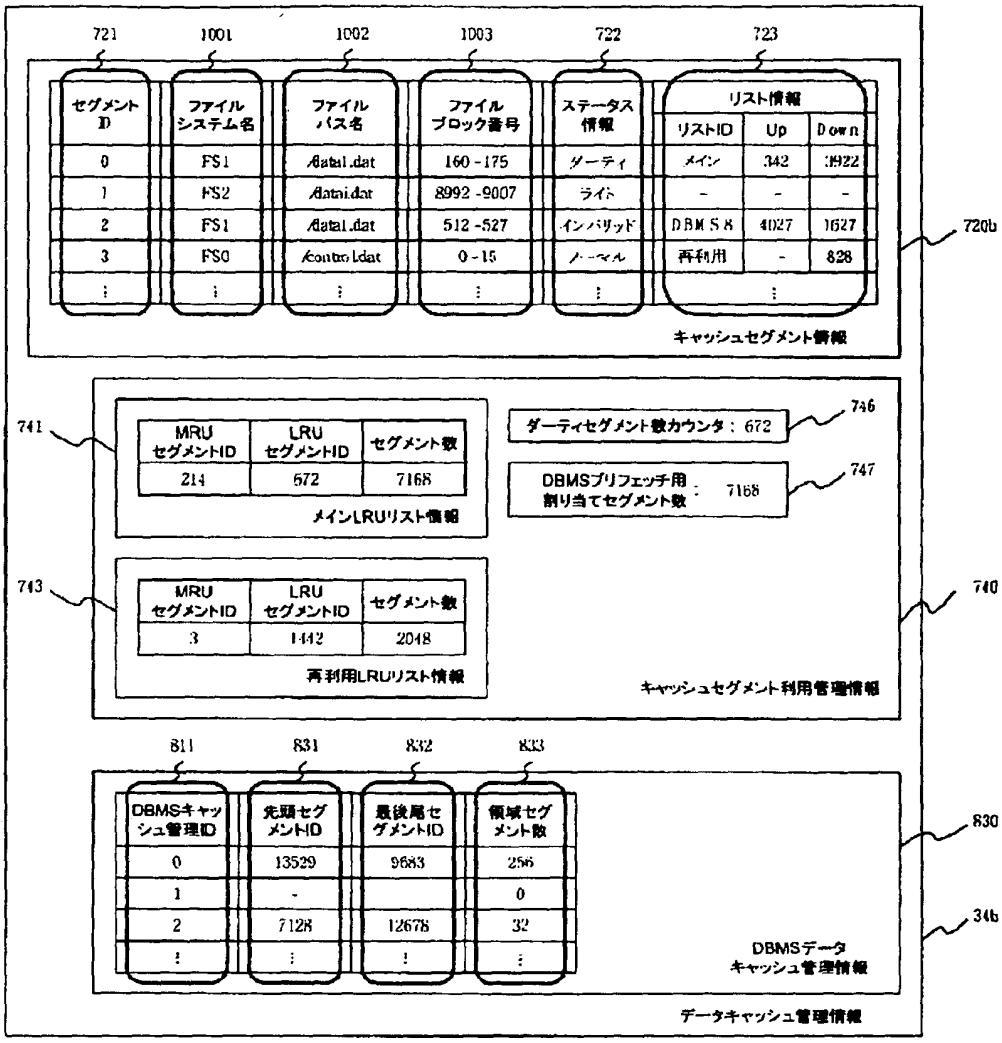
【図２７】

図27


DBMS名
DBMS1
DBMS1
DBMS1
DBMS2
631
データ構造名
T1
T1
T2
T1
661
データ構造ブロック番号
0-4999
5000-9999
0-4999
0-9999
716
ファイルシステム名
FS1
FS1
FS1
FS2
1001
ファイルパス名
/data1.dat
/data1.dat
/data2.dat
/data1.dat
1002
ファイルブロック番号
0-4999
5000-9999
0-4999
0-9999
544
物理記憶装置名
Pdisk 1
Pdisk 0
Pdisk 1
Pdisk 2
514
物理ブロック番号
10240-15239
10240-15239
15240-20239
20180-30179
502
712b
データ構造物理記憶位置情報

【図２８】

図28




フロントページの続き

(72)発明者　喜連川　優
千葉県松戸市二十世紀が丘丸山町17

Ｆターム(参考)　5B005 JJ12 KK02 LL17 MM04 NN22
5B075 NR03 NR20 PQ32 QS07
5B082 FA12